

OM protein - protein search, using sw model

Run on: August 17, 2004, 13:55:45 ; Search time 52.5478 Seconds  
(without alignments)  
1774.398 Million cell updates/sec

Title: US-09-864-675-2  
Perfect score: 1749  
Sequence: 1 MRRDPAPGFSSMLLFGVSLAC.....PGTGVSSSQWSTSPSTLDLN 330

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

		8						
Result		Query						Description
No.	Score	Match	Length	DB	ID			
1	1749	100.0	330	5	AAU11635			Aau11635 Human Neu
2	1749	100.0	422	5	ABB07894			Abb07894 Human neu
3	1722	98.5	330	2	AAW27537			Aaw27537 Rat cereb
4	1720	98.3	426	5	ABB07893			Abb07893 Human neu
5	1575	90.1	860	2	AAW63700			Aaw63700 Receptor
6	1505	86.0	298	5	AAU11636			Aau11636 Human Neu
7	1478	84.5	754	2	AAW27536			Aaw27536 Rat cereb
8	960	54.9	181	2	AAW48380			Aaw48380 Mus muscu
9	960	54.9	181	6	ABG71637			Abg71637 Murine se

10	881	50.4	469	6	ABG71639	Abg71639	Human sec
11	881	50.4	647	2	AAW48383	Aaw48383	Homo sapi
12	881	50.4	647	6	ABG71644	Abg71644	Human thi
13	875	50.0	469	2	AAW48382	Aaw48382	Homo sapi
14	842	48.1	407	2	AAW48381	Aaw48381	Homo sapi
15	842	48.1	407	6	ABG71638	Abg71638	Human mem
16	821	46.9	605	6	ABG71636	Abg71636	Murine me
17	816	46.7	605	2	AAW48379	Aaw48379	Mus muscu
18	748	42.8	139	2	AAW48388	Aaw48388	Undefined
19	748	42.8	139	6	ABG71645	Abg71645	Don-1 ass
20	688	39.3	182	2	AAW27538	Aaw27538	Human cer
21	524	30.0	422	4	AAG67901	Aag67901	Human neu
22	524	30.0	422	4	AAG67939	Aag67939	Human neu
23	523	29.9	422	2	AAR67258	Aar67258	Human gli
24	523	29.9	422	2	AAR96081	Aar96081	Glial gro
25	523	29.9	422	2	AAW09372	Aaw09372	Human GGF
26	523	29.9	422	2	AAW09371	Aaw09371	Human neu
27	523	29.9	422	2	AAR86628	Aar86628	Mature hG
28	523	29.9	422	2	AAR87466	Aar87466	Glial gro
29	522	29.8	422	2	AAR55654	Aar55654	GGF-II en
30	522	29.8	422	2	AAR46923	Aar46923	GGF-II en
31	521	29.8	418	5	ABJ00011	Abj00011	Human neu
32	521	29.8	418	5	ABJ00049	Abj00049	Human neu
33	508	29.0	422	2	AAR87467	Aar87467	Glial gro
34	485	27.7	782	4	AAB67751	Aab67751	Amino aci
35	474	27.1	855	4	AAB67757	Aab67757	Amino aci
36	460	26.3	342	4	AAB67754	Aab67754	Amino aci
37	452	25.8	323	4	AAB67753	Aab67753	Amino aci
38	450	25.7	317	4	AAB67752	Aab67752	Amino aci
39	354	20.2	204	4	AAG67902	Aag67902	Human neu
40	354	20.2	204	4	AAG67940	Aag67940	Human neu
41	354	20.2	204	5	ABJ00012	Abj00012	Human neu
42	354	20.2	204	5	ABJ00050	Abj00050	Human neu
43	353.5	20.2	263	2	AAW09360	Aaw09360	Bovine ne
44	353	20.2	280	2	AAR67244	Aar67244	Bovine gl
45	353	20.2	280	2	AAR98762	Aar98762	Glial gro

# ALIGNMENTS

## RESULT 1

AAU11635

ID AAU11635 standard; protein; 330 AA.

XX

AC AAU11635;

XX

DT 12-MAR-2002 (first entry)

XX

DE Human Neuregulin-2alpha, NRG-2alpha.

XX

KW Human; neuregulin-2; NRG-2alpha; NRG-2beta; mitogenesis; cell survival;

KW cell growth; cell differentiation; erbB receptor; cardiomyopathy;

KW ischaemic damage; cardiac trauma; heart failure; atherosclerosis;

KW vascular lesion; vascular hypertension;

KW degenerative congenital vascular disease; myasthenia gravis;

KW neurodegenerative disorder; peripheral neuropathy;

KW sensory nerve fiber neuropathy; motor fiber neuropathy;  
 KW sensory nerve fiber neuropathy; multiple sclerosis;  
 KW amyotrophic lateral sclerosis; spinal muscular atrophy; nerve injury;  
 KW Alzheimer's disease; Parkinson's disease; cerebellar ataxia;  
 KW spinal cord injury; tumour; neurofibromatosis; transgenic animal.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200189568-A1.  
 XX  
 PD 29-NOV-2001.  
 XX  
 PF 23-MAY-2001; 2001WO-US016896.  
 XX  
 PR 23-MAY-2000; 2000US-0206495P.  
 XX  
 PA (CENE-) CENES PHARM INC.  
 XX  
 PI Marchionni MA;  
 XX  
 DR WPI; 2002-097612/13.  
 DR N-PSDB; AAS18019.  
 XX  
 PT Neuregulin-2 polypeptide and polynucleotide useful for treating multiple  
 PT sclerosis, spinal muscular atrophy, nerve injury, Alzheimer's disease, by  
 PT increasing mitogenesis, survival, growth or differentiation of a cell.  
 XX  
 PS Claim 53; Fig 7; 79pp; English.  
 XX  
 CC The invention relates to a substantially pure neuregulin (NRG)-2  
 CC polypeptide comprising or consisting of a sequence for human NRG-2alpha  
 CC or NRG-2beta (clone 2b7) and the polynucleotides encoding the. Also  
 CC included are a vector expressing the protein, a host cell comprising the  
 CC vector, a transgenic non-human animal transformed with the vector or  
 CC having a knockout mutation in one or both NRG-2 alleles and an anti-NRG-2  
 CC antibody. Analysis of mutations in NRG-2 in an individual is useful for  
 CC diagnosing an increased likelihood of developing a NRG-2-related disease  
 CC or condition in a test subject. NRG-2 is useful for increasing the  
 CC mitogenesis, survival, growth or differentiation of a cell (e.g. a  
 CC neuronal cell), where the cell expresses an erbB receptor. NRG-2 is  
 CC useful for treating diseases and disorders such as cardiomyopathy  
 CC (preferably degenerative congenital disease), ischaemic damage, cardiac  
 CC trauma or heart failure or which has a condition affecting smooth muscle  
 CC which include atherosclerosis, vascular lesion, vascular hypertension,  
 CC and degenerative congenital vascular disease, myasthenia gravis, a  
 CC neurodegenerative disorder, peripheral neuropathy, a sensory nerve fiber  
 CC neuropathy, a motor fiber and a sensory nerve fiber neuropathy, multiple  
 CC sclerosis, amyotrophic lateral sclerosis, spinal muscular atrophy, nerve  
 CC injury, Alzheimer's disease, Parkinson's disease, cerebellar ataxia, and  
 CC spinal cord injury. The antibody is useful for treatment of a tumour  
 CC comprising inhibiting proliferation of a tumour cell preferably a glial  
 CC tumour cell, for treating of neurofibromatosis by inhibiting glial cell  
 CC mitogenesis. The present sequence represents NRG-2alpha  
 XX  
 SQ Sequence 330 AA;

Query Match

100.0%; Score 1749; DB 5; Length 330;

Best Local Similarity 100.0%; Pred. No. 1.1e-108;  
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60

Qy     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAF 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAF 120

Qy    121 PLDTNGKNLKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWF 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 PLDTNGKNLKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWF 180

Qy    181 DGKELNRSRDIRIKYGNRKNRSLQFNKVVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 DGKELNRSRDIRIKYGNRKNRSLQFNKVVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240

Qy    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLY 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLY 300

Qy    301 DPKQSVLWDTPGTGVSSSQWSTSPSTLDLN 330
        ||||||||||||||||||||||||||||||||||||
Db    301 DPKQSVLWDTPGTGVSSSQWSTSPSTLDLN 330

```

# RESULT 2

ABB07894

ID ABB07894 standard; protein; 422 AA.

XX

AC ABB07894;

XX

DT 03-JUL-2002 (first entry)

XX

DE Human neuregulin 2 isoform 6.

XX

KW Human; MUC1; mucin; glycoprotein; cytostatic; cancer; tumour; ECD;  
KW extracellular domain; neuregulin 2; isoform.

XX

OS Homo sapiens.

XX

PN WO200222685-A2.

XX

PD 21-MAR-2002.

XX

PF 11-SEP-2001; 2001WO-US028548.

XX

PR 11-SEP-2000; 2000US-0231841P.

XX

PA (KUFE/) KUFE D W.

PA (OHNO/) OHNO T.

XX

PI Kufe DW, Ohno T;

XX

DR WPI; 2002-339864/37.

XX  
PT Use of a mucin glycoprotein (MUC1) extracellular domain antagonist for  
PT manufacturing a medicant that inhibits the proliferation of MUC-1  
PT expressing cancer cells and that can treat cancers and reduce tumor  
PT growth.  
XX  
PS Claim 6; Page 56-58; 74pp; English.  
XX  
CC The invention relates to the use of a MUC1 (mucin glycoprotein)  
CC extracellular domain (ECD) antagonist for the manufacture of a medicant  
CC to inhibit the proliferation of MUC-1 expressing cancer cells. MUC1 ECD  
CC antagonists (optionally combined with a pharmaceutical carrier) can be  
CC administered to inhibit proliferation of MUC1-expressing cancer cells,  
CC useful to treat cancers e.g. skin cancer, prostate cancer and leukemia,  
CC especially in humans. The method may also be combined with administration  
CC of a chemotherapeutic agent (e.g. an alkylating agent, topoisomerase etc)  
CC or radiation to treat cancer, especially to reduce tumour growth. The  
CC polypeptides are also useful in screening to identify MUC1 ECD  
CC antagonists. The present sequence represents a human neuregulin 2 isoform  
CC 6, a fragment of which can bind to MUC1/ECD  
XX  
SQ Sequence 422 AA;

Query Match 100.0%; Score 1749; DB 5; Length 422;  
Best Local Similarity 100.0%; Pred. No. 1.5e-108;  
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP	60
Db	93	MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP	152
Qy	61	PASGRVALVKVLDKWPLRSGGLQREQVISVSGCVPLERNQRYIFFLEPTEQPLVFKTAFA	120
Db	153	PASGRVALVKVLDKWPLRSGGLQREQVISVSGCVPLERNQRYIFFLEPTEQPLVFKTAFA	212
Qy	121	PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK	180
Db	213	PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK	272
Qy	181	DGKELNRSRDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS	240
Db	273	DGKELNRSRDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS	332
Qy	241	TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP	300
Db	333	TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP	392
Qy	301	DPKQSVLWDTPGTGVSSSQWSTSPSTLDLN	330
Db	393	DPKQSVLWDTPGTGVSSSQWSTSPSTLDLN	422

RESULT 3  
AAW27537  
ID AAW27537 standard; protein; 330 AA.  
XX  
AC AAW27537;

XX  
 DT 18-DEC-1997 (first entry)  
 XX  
 DE Rat cerebellum derived growth factor 2.  
 XX  
 KW Rat; cerebellum derived growth factor; CDGF2; screening; binding;  
 KW modulation; erbB type receptor; identification; indication; risk;  
 KW proliferation; differentiation; induction; neuron; hyperplasia;  
 KW stem cell culture; intracerebral graft; alleviation; repair;  
 KW behavioural defect; nervous system; central; peripheral; nerve;  
 KW prosthesis; damage; entubulation; cell survival; treatment; injury;  
 KW trauma; ischaemia; ischemia; stroke; infection; disorder; inflammation;  
 KW neurodegeneration; disease; Parkinson's; Huntingdon's;  
 KW amyotrophic lateral sclerosis; sensory; retina;  
 KW spinocerebellar degeneration; multiple sclerosis; neoplasia;  
 KW amalignant glioma; medulloblastoma; neuroectodermal tumour.  
 XX  
 OS Rattus rattus.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1. .23  
 FT /label= sig\_peptide  
 FT Peptide 24. .330  
 FT /label= mat\_peptide  
 FT Region 55  
 FT /note= "potential N-glycosylation site"  
 FT Domain 158. .228  
 FT /label= immunoglobulin\_like\_domain  
 FT Region 186  
 FT /note= "potential N-glycosylation site"  
 FT Domain 252. .297  
 FT /label= epidermal\_growth\_factor\_like\_domain  
 FT Region 253  
 FT /note= "characteristic cysteine of epidermal growth  
 FT factor like domain"  
 FT Region 254  
 FT /note= "potential N-glycosylation site"  
 FT Region 261  
 FT /note= "characteristic cysteine of epidermal growth  
 FT factor like domain"  
 FT Region 267  
 FT /note= "characteristic cysteine of epidermal growth  
 FT factor like domain"  
 FT Region 278  
 FT /note= "characteristic cysteine of epidermal growth  
 FT factor like domain"  
 FT Region 280  
 FT /note= "characteristic cysteine of epidermal growth  
 FT factor like domain"  
 FT Region 289  
 FT /note= "characteristic cysteine of epidermal growth  
 FT factor like domain"  
 XX  
 PN W09709425-A1.  
 XX  
 PD 13-MAR-1997.  
 XX

XX

XX

PA

XX

XX

DR

XX

XX

XX

XX

Best Local Similarity 97.9%; Pred. No. 7.1e-107;

Db

Ov

Db

Ov

121 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGQVGKQSLKCEAAAGNPQPSYRWFK 180

Db 121 PVDPNGKNIKKEVGKILCTDCATRPKLKKMKSQTGEVGEKQSLKCEAAAGNPQPSYRWFK 180  
 QY 181 DGKELNRSRDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240  
 |||||||||||||||||||||||||||||||||||||||||||||:|||||  
 Db 181 DGKELNRSRDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLHVNSVS 240  
 QY 241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP 300  
 |||||||||||||||||||||||||||||||||||||||||||||  
 Db 241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP 300  
 QY 301 DPKQSVLWDTPGTGVSSSQWSTSPSTLDLN 330  
 |||||||||||||||||||||||||||||  
 Db 301 DPKQSVLWDTPGTGVSSSQWSTSPSTLDLN 330

RESULT 4

ABB07893

ID ABB07893 standard; protein; 426 AA.

XX

AC ABB07893;

XX

DT 03-JUL-2002 (first entry)

XX

DE Human neuregulin 2 isoform 5.

XX

KW Human; MUC1; mucin; glycoprotein; cytostatic; cancer; tumour; ECD;  
 KW extracellular domain; neuregulin 2; isoform.

XX

OS Homo sapiens.

XX

PN WO200222685-A2.

XX

PD 21-MAR-2002.

XX

PF 11-SEP-2001; 2001WO-US028548.

XX

PR 11-SEP-2000; 2000US-0231841P.

XX

PA (KUFE/) KUFE D W.

PA (OHNO/) OHNO T.

XX

PI Kufe DW, Ohno T;

XX

DR WPI; 2002-339864/37.

XX

PT Use of a mucin glycoprotein (MUC1) extracellular domain antagonist for  
 PT manufacturing a medicant that inhibits the proliferation of MUC-1  
 PT expressing cancer cells and that can treat cancers and reduce tumor  
 PT growth.

XX

PS Claim 6; Page 53-55; 74pp; English.

XX

CC The invention relates to the use of a MUC1 (mucin glycoprotein)  
 CC extracellular domain (ECD) antagonist for the manufacture of a medicant  
 CC to inhibit the proliferation of MUC-1 expressing cancer cells. MUC1 ECD  
 CC antagonists (optionally combined with a pharmaceutical carrier) can be  
 CC administered to inhibit proliferation of MUC1-expressing cancer cells,



CC useful to treat cancers e.g. skin cancer, prostate cancer and leukemia,  
CC especially in humans. The method may also be combined with administration  
CC of a chemotherapeutic agent (e.g. an alkylating agent, topoisomerase etc)  
CC or radiation to treat cancer, especially to reduce tumour growth. The  
CC polypeptides are also useful in screening to identify MUC1 ECD  
CC antagonists. The present sequence represents a human neuregulin 2 isoform  
CC 5, a fragment of which can bind to MUC1/ECD

XX

SQ Sequence 426 AA;

Query Match 98.3%; Score 1720; DB 5; Length 426;

Best Local Similarity 100.0%; Pred. No. 1.3e-106;

Matches 324; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
          |||
Db      93 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 152

Qy      61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 120
          |||
Db     153 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 212

Qy     121 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKsQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180
          |||
Db     213 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKsQTGQVGEKQSLKCEAAAGNPQPSYRWFK 272

Qy     181 DGKELNRSRDIRIKYGNRKNsRLQFNKVVEDAGEYVCEAENILGKDTVGRLYVNSVS 240
          |||
Db     273 DGKELNRSRDIRIKYGNRKNsRLQFNKVVEDAGEYVCEAENILGKDTVGRLYVNSVS 332

Qy     241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP 300
          |||
Db     333 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP 392

Qy     301 DPKQSVLWDTPGTGVSSSQWSTSP 324
          |||
Db     393 DPKQSVLWDTPGTGVSSSQWSTSP 416
```

#### RESULT 5

AAW63700

ID AAW63700 standard; protein; 860 AA.

XX

AC AAW63700;

XX

DT 29-SEP-1998 (first entry)

XX

DE Receptor type tyrosine kinase ErbB ligand.

XX

KW Receptor type tyrosine kinase ErbB; ligand; diagnostic agent;  
KW nervous disease; cancer.

XX

OS Rattus sp.

XX

PN JP10179166-A.

XX

PD 07-JUL-1998.

XX  
 PF 25-DEC-1996; 96JP-00356998.  
 XX  
 PR 25-DEC-1996; 96JP-00356998.  
 XX  
 PA (HIGA/) HIGASHIYAMA S.  
 XX  
 DR WPI; 1998-430952/37.  
 DR N-PSDB; AAV43674.  
 XX  
 PT Gene coding the ligand of the tyrosine kinase ErbB receptor - useful for  
 PT diagnosing and treating nervous diseases and cancer.  
 XX  
 PS Claim 1; Page 9-13; 17pp; Japanese.  
 XX  
 CC This represents the ligand of receptor type tyrosine kinase ErbB. A  
 CC prokaryotic or eukaryotic host cell transformed by a recombinant vector  
 CC containing the encoding DNA can be used for the recombinant production of  
 CC the protein. The invention provides a method for inhibiting the formation  
 CC of the ligand of receptor type tyrosine kinase ErbB in an animal using an  
 CC antibody recognizing the protein. The ligand of the tyrosine kinase ErbB  
 CC receptor and associated materials can be used for treating or diagnosing  
 CC nervous diseases and cancers  
 XX  
 SQ Sequence 860 AA;

Query Match 90.1%; Score 1575; DB 2; Length 860;  
 Best Local Similarity 97.4%; Pred. No. 1.3e-96;  
 Matches 296; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60  
 ||||| ||||||||||||||||||||||||||||||||||||||||  
 Db 109 MRRDPAPGSSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLAPAGGSSSNSTREP 168  
 Qy 61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAF 120  
 ||||| ||||||||||||||||||||||||||||||||||||||||  
 Db 169 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCAPLERNQRYIFFLEPTEQPLVFKTAF 228  
 Qy 121 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWF 180  
 |:| ||||:||||||||||||||||||||||:||||||||||||||  
 Db 229 PVDPNGKNIKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWF 288  
 Qy 181 DGKELNRSRDIRIKYGNRKNLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240  
 ||||| ||||||||||||||||||||||||||||||||||||||||  
 Db 289 DGKELNRSRDIRIKYGNRKNLQFNKVKVEDAGEYVCEAENILGKDTVGRGLHVNSVS 348  
 Qy 241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP 300  
 ||||| ||||||||||||||||||||||||||||||||||||||||  
 Db 349 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP 408  
 Qy 301 DPKQ 304  
 ||||  
 Db 409 DPKQ 412

RESULT 6  
 AAU11636

ID AAU11636 standard; protein; 298 AA.  
XX  
AC AAU11636;  
XX  
DT 12-MAR-2002 (first entry)  
XX  
DE Human Neuregulin-2beta, NRG-2beta.  
XX  
KW Human; neuregulin-2; NRG-2alpha; NRG-2beta; mitogenesis; cell survival;  
KW cell growth; cell differentiation; erbB receptor; cardiomyopathy;  
KW ischaemic damage; cardiac trauma; heart failure; atherosclerosis;  
KW vascular lesion; vascular hypertension;  
KW degenerative congenital vascular disease; myasthenia gravis;  
KW neurodegenerative disorder; peripheral neuropathy;  
KW sensory nerve fiber neuropathy; motor fiber neuropathy;  
KW sensory nerve fiber neuropathy; multiple sclerosis;  
KW amyotrophic lateral sclerosis; spinal muscular atrophy; nerve injury;  
KW Alzheimer's disease; Parkinson's disease; cerebellar ataxia;  
KW spinal cord injury; tumour; neurofibromatosis; transgenic animal.  
XX  
OS Homo sapiens.  
XX  
PN WO200189568-A1.  
XX  
PD 29-NOV-2001.  
XX  
PF 23-MAY-2001; 2001WO-US016896.  
XX  
PR 23-MAY-2000; 2000US-0206495P.  
XX  
PA (CENE-) CENES PHARM INC.  
XX  
PI Marchionni MA;  
XX  
DR WPI; 2002-097612/13.  
DR N-PSDB; AAS18020.  
XX  
PT Neuregulin-2 polypeptide and polynucleotide useful for treating multiple  
PT sclerosis, spinal muscular atrophy, nerve injury, Alzheimer's disease, by  
PT increasing mitogenesis, survival, growth or differentiation of a cell.  
XX  
PS Claim 53; Fig 9; 79pp; English.  
XX  
CC The invention relates to a substantially pure neuregulin (NRG)-2  
CC polypeptide comprising or consisting of a sequence for human NRG-2alpha  
CC or NRG-2beta (clone 2b7) and the polynucleotides encoding the. Also  
CC included are a vector expressing the protein, a host cell comprising the  
CC vector, a transgenic non-human animal transformed with the vector or  
CC having a knockout mutation in one or both NRG-2 alleles and an anti-NRG-2  
CC antibody. Analysis of mutations in NRG-2 in an individual is useful for  
CC diagnosing an increased likelihood of developing a NRG-2-related disease  
CC or condition in a test subject. NRG-2 is useful for increasing the  
CC mitogenesis, survival, growth or differentiation of a cell (e.g. a  
CC neuronal cell), where the cell expresses an erbB receptor. NRG-2 is  
CC useful for treating diseases and disorders such as cardiomyopathy  
CC (preferably degenerative congenital disease), ischaemic damage, cardiac  
CC trauma or heart failure or which has a condition affecting smooth muscle

CC which include atherosclerosis, vascular lesion, vascular hypertension,  
 CC and degenerative congenital vascular disease, myasthenia gravis, a  
 CC neurodegenerative disorder, peripheral neuropathy, a sensory nerve fiber  
 CC neuropathy, a motor fiber and a sensory nerve fiber neuropathy, multiple  
 CC sclerosis, amyotrophic lateral sclerosis, spinal muscular atrophy, nerve  
 CC injury, Alzheimer's disease, Parkinson's disease, cerebellar ataxia, and  
 CC spinal cord injury. The antibody is useful for treatment of a tumour  
 CC comprising inhibiting proliferation of a tumour cell preferably a glial  
 CC tumour cell, for treating of neurofibromatosis by inhibiting glial cell  
 CC mitogenesis. The present sequence represents NRG-2beta  
 XX  
 SQ Sequence 298 AA;

Query Match 86.0%; Score 1505; DB 5; Length 298;  
 Best Local Similarity 98.6%; Pred. No. 1.8e-92;  
 Matches 285; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

Qy      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
      |||
Db      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60

Qy      61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 120
      |||
Db      61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 120

Qy      121 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180
      |||
Db      121 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180

Qy      181 DGKELNRSRDIRIKYGNRKNRSLQFNKVVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240
      |||
Db      181 DGKELNRSRDIRIKYGNRKNRSLQFNKVVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240

Qy      241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRC 289
      |||
Db      241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDR 289
  
```

# RESULT 7

AAW27536

ID AAW27536 standard; protein; 754 AA.

XX

AC AAW27536;

XX

DT 18-DEC-1997 (first entry)

XX

DE Rat cerebellum derived growth factor 1.

XX

KW Rat; cerebellum derived growth factor; CDGF1; screening; binding;

KW modulation; erbB type receptor; identification; indication; risk;

KW proliferation; differentiation; induction; neuron; hyperplasia;

KW stem cell culture; intracerebral graft; alleviation; repair;

KW behavioural defect; nervous system; central; peripheral; nerve;

KW prosthesis; damage; entubulation; cell survival; treatment; injury;

KW trauma; ischaemia; ischemia; stroke; infection; disorder; inflammation;

KW neurodegeneration; disease; Parkinson's; Huntingdon's;

KW amyotrophic lateral sclerosis; sensory; retina;

KW spinocerebellar degeneration; multiple sclerosis; neoplasia;  
 KW amalignant glioma; medulloblastoma; neuroectodermal tumour.  
 XX  
 OS Rattus rattus.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1. .23  
 FT /label= sig\_peptide  
 FT Peptide 24. .754  
 FT /label= mat\_peptide  
 FT Region 55  
 FT /note= "potential N-glycosylation site"  
 FT Domain 158. .228  
 FT /label= immunoglobulin\_like\_domain  
 FT Region 186  
 FT /note= "potential N-glycosylation site"  
 FT Domain 252. .297  
 FT /label= epidermal\_growth\_factor\_like\_domain  
 FT Region 253  
 FT /note= "characteristic cysteine of epidermal growth  
 FT factor like domain"  
 FT Region 254  
 FT /note= "potential N-glycosylation site"  
 FT Region 261  
 FT /note= "characteristic cysteine of epidermal growth  
 FT factor like domain"  
 FT Region 267  
 FT /note= "characteristic cysteine of epidermal growth  
 FT factor like domain"  
 FT Region 278  
 FT /note= "characteristic cysteine of epidermal growth  
 FT factor like domain"  
 FT Region 280  
 FT /note= "characteristic cysteine of epidermal growth  
 FT factor like domain"  
 FT Region 289  
 FT /note= "characteristic cysteine of epidermal growth  
 FT factor like domain"  
 FT Region 296  
 FT /note= "potential N-glycosylation site"  
 FT Cleavage-site 314. .315  
 FT /label= potential\_proteolytic\_site  
 FT Domain 316. .338  
 FT /label= putative\_transmembrane\_domain  
 XX  
 PN WO9709425-A1.  
 XX  
 PD 13-MAR-1997.  
 XX  
 PF 09-SEP-1996; 96WO-US014484.  
 XX  
 PR 08-SEP-1995; 95US-00525864.  
 XX  
 PA (HARD ) HARVARD COLLEGE.  
 PA (STRD ) UNIV LELAND S STANFORD.  
 XX  
 PI Chang H;



## RESULT 8

AAW48380

ID AAW48380 standard; protein; 181 AA.

XX

AC AAW48380;

XX

DT 17-AUG-1998 (first entry)

XX

DE Mus musculus don-1 polypeptide.

XX

KW Murine; don-1 gene; melanoma; treatment; adenocarcinoma; epithelial cell;

KW proliferation; stimulation; treatment; tumours; skin; oesophagus; lung;

KW breast; liver; pancreas; colon; prostate; gastrointestinal tract; uterus;

KW wound healing; transmembrane.

XX

OS Mus musculus.

XX

FH Key Location/Qualifiers

FT Domain 104. .140

FT /note= "EGF domain"

XX

PN WO9807736-A1.

XX

PD 26-FEB-1998.

XX

PF 18-AUG-1997; 97WO-US014585.

XX

PR 19-AUG-1996; 96US-00699591.

PR 19-NOV-1996; 96US-00753007.

XX

PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.

XX

PI Gearing DP, Busfield SJ;

XX

DR WPI; 1998-169084/15.

DR N-PSDB; AAV17813.

XX

PT Mouse and human don-1 polypeptide(s) - useful for treatment of melanomas  
PT and adenocarcinoma(s), and for wound healing.

XX

PS Claim 25; Fig 2; 121pp; English.

XX

CC The sequence is that encoded by a murine don-1 gene splice variant. Don-1  
CC polypeptides stimulate proliferation of epithelial cells and thus are  
CC implicated in melanomas and adenocarcinomas in which epithelial cells  
CC proliferate out of control. Compounds that interfere with don-1 mediated  
CC cell proliferation can be used in the treatment of tumours such as  
CC melanomas and adenocarcinomas of the skin, oesophagus, lung, breast,  
CC liver, pancreas, gastrointestinal tract, colon, prostate or uterus.  
CC Alternatively, don-1 polypeptides can be used to stimulate epithelial  
CC cell proliferation, e.g. for wound healing

XX

SQ Sequence 181 AA;

Query Match 54.9%; Score 960; DB 2; Length 181;  
Best Local Similarity 97.8%; Pred. No. 2.2e-56;  
Matches 177; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```
Qy      150 MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGGKELNRSRDIRIKYGNRKN SRLQFNKV 209
          |||||:|||||
Db      1   MKSQTGEVGEKQSLKCEAAAGNPQPSYRWFKDGGKELNRSRDIRIKYGNVRKN SRLQFNKV 60

Qy      210 KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 269
          :|||||
Db      61 RVEDAGEYVCEAENILGKDTVGRGLHVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 120

Qy      270 IEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQSVLWDTPGTGVSSSQWSTSPSTLDL 329
          |||||
Db      121 IEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQSVLWDTPGTGVSSSQWSTSPSTLDL 180

Qy      330 N 330
          |
Db      181 N 181
```

RESULT 9

ABG71637

ID ABG71637 standard; protein; 181 AA.

XX

AC ABG71637;

XX

DT 14-JAN-2003 (first entry)

XX

DE Murine secreted splice variant of Don-1.

XX

KW Murine; Don-1; epidermal growth factor; EGF; neuregulin; mouse;  
KW glycoprotein ligand; cell proliferation; cell proliferative disorder;  
KW carcinoma; adenocarcinoma cell; myeloma; cell differentiation;  
KW cell survival; epithelial cell; wound healing; tumour formation; brain;  
KW vulnerary; cytostatic.

XX

OS Mus sp.

XX

PN US2002127594-A1.

XX

PD 12-SEP-2002.

XX

PF 12-MAR-2002; 2002US-00096241.

XX

PR 22-JUN-2000; 2000US-00599789.

XX

PA (GEAR/) GEARING D P.

PA (BUSF/) BUSFIELD S J.

XX

PI Gearing DP, Busfield SJ;

XX

DR WPI; 2003-039584/03.

DR N-PSDB; ABS56034.

XX

PT Novel Don-1 polypeptide useful for stimulating proliferation of cells,  
PT for identifying proteins that interact with Don-1, and for regulating





KW carcinoma; adenocarcinoma cell; myeloma; cell differentiation;  
KW cell survival; epithelial cell; wound healing; tumour formation; brain;  
KW vulnerary; cytostatic.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Misc-difference 14

FT /note= "Encoded by AA"

XX

PN US2002127594-A1.

XX

PD 12-SEP-2002.

XX

PF 12-MAR-2002; 2002US-00096241.

XX

PR 22-JUN-2000; 2000US-00599789.

XX

PA (GEAR/) GEARING D P.

PA (BUSF/) BUSFIELD S J.

XX

PI Gearing DP, Busfield SJ;

XX

DR WPI; 2003-039584/03.

DR N-PSDB; ABS56036.

XX

PT Novel Don-1 polypeptide useful for stimulating proliferation of cells,  
PT for identifying proteins that interact with Don-1, and for regulating  
PT tumor formation and progression in brain.

XX

PS Claim 25; Fig 4; 66pp; English.

XX

CC The present invention relates to the isolation of a novel gene called Don  
CC -1, and alternate splice variants of Don-1, which are related to  
CC epidermal growth factors (EGF) such as neuregulins. Don-1 polypeptides  
CC are glycoprotein ligands. Both murine and human Don-1 sequences are  
CC cloned. The mouse Don-1 gene maps to chromosome 18. Don-1 polypeptides  
CC are useful for stimulating proliferation of a cell. Antibodies to Don-1  
CC polypeptides are useful for detecting Don-1 in a sample. The Don-1  
CC polypeptides are useful for treating and diagnosing cell proliferative  
CC disorders and play a role in the proliferation of carcinomas e.g.  
CC adenocarcinoma, myeloma, in cell differentiation, proliferation and  
CC survival. The polypeptides are also useful for inhibiting proliferation  
CC of adenocarcinoma cells, for stimulating the proliferation of cells such  
CC as epithelial cells to promote wound healing, for identifying proteins  
CC that interact with Don-1, and for regulating tumour formation and  
CC progression in the brain. The polynucleotide sequences encoding Don-1 may  
CC be used in gene therapy. The present sequence represents human second  
CC splice variant of Don-1

XX

SQ Sequence 469 AA;

Query Match 50.4%; Score 881; DB 6; Length 469;

Best Local Similarity 100.0%; Pred. No. 1.2e-50;

Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 ATRPKLKKMKSQTGQVGKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN 201

```

Db          |||||
31 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN 90

Qy          202 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC 261
          |||||
Db          91 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC 150

Qy          262 VNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ 304
          |||||
Db          151 VNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ 193

```

RESULT 11

AAW48383

ID AAW48383 standard; protein; 647 AA.

XX

AC AAW48383;

XX

DT 17-AUG-1998 (first entry)

XX

DE Homo sapiens don-1 polypeptide.

XX

KW Murine; don-1 gene; melanoma; treatment; adenocarcinoma; epithelial cell;  
 KW proliferation; stimulation; treatment; tumours; skin; oesophagus; lung;  
 KW breast; liver; pancreas; colon; prostate; gastrointestinal tract; uterus;  
 KW wound healing; transmembrane.

XX

OS Homo sapiens.

XX

FH	Key	Location/Qualifiers
----	-----	---------------------

FT	Domain	54. .108
----	--------	----------

FT		/note= "Ig domain"
----	--	--------------------

FT	Domain	142. .178
----	--------	-----------

FT		/note= "EGF domain"
----	--	---------------------

FT	Domain	203. .225
----	--------	-----------

FT		/note= "transmembrane domain"
----	--	-------------------------------

FT	Domain	226. .647
----	--------	-----------

FT		/note= "cytoplasmic domain"
----	--	-----------------------------

XX

PN WO9807736-A1.

XX

PD 26-FEB-1998.

XX

PF 18-AUG-1997; 97WO-US014585.

XX

PR 19-AUG-1996; 96US-00699591.

PR 19-NOV-1996; 96US-00753007.

XX

PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.

XX

PI Gearing DP, Busfield SJ;

XX

DR WPI; 1998-169084/15.

DR N-PSDB; AAV17816.

XX

PT Mouse and human don-1 polypeptide(s) - useful for treatment of melanomas  
 PT and adenocarcinoma(s), and for wound healing.

XX  
 PS Claim 25; Fig 7; 121pp; English.  
 XX  
 CC The sequence is that encoded by a human don-1 gene splice variant. Don-1  
 CC polypeptides stimulate proliferation of epithelial cells and thus are  
 CC implicated in melanomas and adenocarcinomas in which epithelial cells  
 CC proliferate out of control. Compounds that interfere with don-1 mediated  
 CC cell proliferation can be used in the treatment of tumours such as  
 CC melanomas and adenocarcinomas of the skin, oesophagus, lung, breast,  
 CC liver, pancreas, gastrointestinal tract, colon, prostate or uterus.  
 CC Alternatively, don-1 polypeptides can be used to stimulate epithelial  
 CC cell proliferation, e.g. for wound healing  
 XX  
 SQ Sequence 647 AA;

Query Match 50.4%; Score 881; DB 2; Length 647;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-50;  
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN 201  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 31 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN 90  
 Qy 202 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRLYVNSVSTTLSSWSGHARKCNETAKSYC 261  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 91 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRLYVNSVSTTLSSWSGHARKCNETAKSYC 150  
 Qy 262 VNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ 304  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 151 VNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ 193

RESULT 12

ABG71644

ID ABG71644 standard; protein; 647 AA.

XX

AC ABG71644;

XX

DT 14-JAN-2003 (first entry)

XX

DE Human third splice variant of Don-1.

XX

KW Human; Don-1; epidermal growth factor; EGF; neuregulin;  
 KW glycoprotein ligand; cell proliferation; cell proliferative disorder;  
 KW carcinoma; adenocarcinoma cell; myeloma; cell differentiation;  
 KW cell survival; epithelial cell; wound healing; tumour formation; brain;  
 KW vulnerary; cytostatic.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Misc-difference 14

FT /note= "Encoded by AA"

FT Misc-difference 310

FT /note= "Encoded by AGC"

XX

PN US2002127594-A1.

XX  
PD 12-SEP-2002.  
XX  
PF 12-MAR-2002; 2002US-00096241.  
XX  
PR 22-JUN-2000; 2000US-00599789.  
XX  
PA (GEAR/) GEARING D P.  
PA (BUSF/) BUSFIELD S J.  
XX  
PI Gearing DP, Busfield SJ;  
XX  
DR WPI; 2003-039584/03.  
DR N-PSDB; ABS56045.  
XX  
PT Novel Don-1 polypeptide useful for stimulating proliferation of cells,  
PT for identifying proteins that interact with Don-1, and for regulating  
PT tumor formation and progression in brain.  
XX  
PS Claim 25; Fig 7; 66pp; English.  
XX  
CC The present invention relates to the isolation of a novel gene called Don  
CC -1, and alternate splice variants of Don-1, which are related to  
CC epidermal growth factors (EGF) such as neuregulins. Don-1 polypeptides  
CC are glycoprotein ligands. Both murine and human Don-1 sequences are  
CC cloned. The mouse Don-1 gene maps to chromosome 18. Don-1 polypeptides  
CC are useful for stimulating proliferation of a cell. Antibodies to Don-1  
CC polypeptides are useful for detecting Don-1 in a sample. The Don-1  
CC polypeptides are useful for treating and diagnosing cell proliferative  
CC disorders and play a role in the proliferation of carcinomas e.g.  
CC adenocarcinoma, myeloma, in cell differentiation, proliferation and  
CC survival. The polypeptides are also useful for inhibiting proliferation  
CC of adenocarcinoma cells, for stimulating the proliferation of cells such  
CC as epithelial cells to promote wound healing, for identifying proteins  
CC that interact with Don-1, and for regulating tumour formation and  
CC progression in the brain. The polynucleotide sequences encoding Don-1 may  
CC be used in gene therapy. The present sequence represents human third  
CC splice variant of Don-1  
XX  
SQ Sequence 647 AA;

Query Match 50.4%; Score 881; DB 6; Length 647;  
Best Local Similarity 100.0%; Pred. No. 1.7e-50;  
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	142	ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN	201
Db	31	ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN	90
Qy	202	SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLVNSVSTTLSSWSGHARKCNETAKSYC	261
Db	91	SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLVNSVSTTLSSWSGHARKCNETAKSYC	150
Qy	262	VNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ	304
Db	151	VNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ	193

RESULT 13

AAW48382

ID AAW48382 standard; protein; 469 AA.

XX

AC AAW48382;

XX

DT 17-AUG-1998 (first entry)

XX

DE Homo sapiens don-1 polypeptide.

XX

KW Murine; don-1 gene; melanoma; treatment; adenocarcinoma; epithelial cell;  
 KW proliferation; stimulation; treatment; tumours; skin; oesophagus; lung;  
 KW breast; liver; pancreas; colon; prostate; gastrointestinal tract; uterus;  
 KW wound healing; transmembrane.

XX

OS Homo sapiens.

XX

FH	Key	Location/Qualifiers
FT	Domain	54. .108
FT		/note= "Ig domain"
FT	Domain	142. .178
FT		/note= "EGF domain"
FT	Domain	203. .225
FT		/note= "transmembrane domain"
FT	Domain	226. .469
FT		/note= "cytoplasmic domain"

XX

PN WO9807736-A1.

XX

PD 26-FEB-1998.

XX

PF 18-AUG-1997; 97WO-US014585.

XX

PR 19-AUG-1996; 96US-00699591.

PR 19-NOV-1996; 96US-00753007.

XX

PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.

XX

PI Gearing DP, Busfield SJ;

XX

DR WPI; 1998-169084/15.

DR N-PSDB; AAV17815.

XX

PT Mouse and human don-1 polypeptide(s) - useful for treatment of melanomas  
 PT and adenocarcinoma(s), and for wound healing.

XX

PS Claim 25; Fig 4; 121pp; English.

XX

CC The sequence is that encoded by a human don-1 gene splice variant. Don-1  
 CC polypeptides stimulate proliferation of epithelial cells and thus are  
 CC implicated in melanomas and adenocarcinomas in which epithelial cells  
 CC proliferate out of control. Compounds that interfere with don-1 mediated  
 CC cell proliferation can be used in the treatment of tumours such as  
 CC melanomas and adenocarcinomas of the skin, oesophagus, lung, breast,  
 CC liver, pancreas, gastrointestinal tract, colon, prostate or uterus.  
 CC Alternatively, don-1 polypeptides can be used to stimulate epithelial

CC cell proliferation, e.g. for wound healing  
XX  
SQ Sequence 469 AA;

Query Match 50.0%; Score 875; DB 2; Length 469;  
Best Local Similarity 99.4%; Pred. No. 3e-50;  
Matches 162; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      142 ATRPKLKKMKSQTGQVGEEKQLKCEAAAGNPQPSYRWFKDGGKELNRSRDIRIKYGNRKN 201
          |||
Db      31 ATRPKLKKMKSQTGQVGEEKQLKCEAAAGNPQPSYRWFKDGGKELNRSRDIRIKYGNRKN 90

Qy      202 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRLYVNSVSTTLSSWSGHARKCNETAKSYC 261
          |||
Db      91 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRLYVNSVSTTLSSWSGHARKCNETAKSYC 150

Qy      262 VNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ 304
          |||
Db      151 VNGGVCYYIEGINQLSCKCPNGFFAQRCLLEKLPLRLYMPDPKQ 193
```

RESULT 14

AAW48381

ID AAW48381 standard; protein; 407 AA.

XX

AC AAW48381;

XX

DT 17-AUG-1998 (first entry)

XX

DE Homo sapiens don-1 polypeptide.

XX

KW Murine; don-1 gene; melanoma; treatment; adenocarcinoma; epithelial cell;  
KW proliferation; stimulation; treatment; tumours; skin; oesophagus; lung;  
KW breast; liver; pancreas; colon; prostate; gastrointestinal tract; uterus;  
KW wound healing; transmembrane.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Domain 16. .70

FT /note= "Ig domain"

FT Domain 104. .140

FT /note= "EGF domain"

FT Region 157. .164

FT /note= "juxtamembrane region"

FT Domain 173. .195

FT /note= "transmembrane domain"

FT Domain 196. .407

FT /note= "cytoplasmic domain"

XX

PN WO9807736-A1.

XX

PD 26-FEB-1998.

XX

PF 18-AUG-1997; 97WO-US014585.

XX

PR 19-AUG-1996; 96US-00699591.

PR 19-NOV-1996; 96US-00753007.  
 XX  
 PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.  
 XX  
 PI Gearing DP, Busfield SJ;  
 XX  
 DR WPI; 1998-169084/15.  
 DR N-PSDB; AAV17814.  
 XX  
 PT Mouse and human don-1 polypeptide(s) - useful for treatment of melanomas  
 PT and adenocarcinoma(s), and for wound healing.  
 XX  
 PS Claim 25; Fig 3; 121pp; English.  
 XX  
 CC The sequence is that encoded by a human don-1 gene splice variant. Don-1  
 CC polypeptides stimulate proliferation of epithelial cells and thus are  
 CC implicated in melanomas and adenocarcinomas in which epithelial cells  
 CC proliferate out of control. Compounds that interfere with don-1 mediated  
 CC cell proliferation can be used in the treatment of tumours such as  
 CC melanomas and adenocarcinomas of the skin, oesophagus, lung, breast,  
 CC liver, pancreas, gastrointestinal tract, colon, prostate or uterus.  
 CC Alternatively, don-1 polypeptides can be used to stimulate epithelial  
 CC cell proliferation, e.g. for wound healing  
 XX  
 SQ Sequence 407 AA;

Query Match 48.1%; Score 842; DB 2; Length 407;  
 Best Local Similarity 98.7%; Pred. No. 4e-48;  
 Matches 156; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 150 MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGGKELNRSRDIRIKYGNRKN SRLQFNKV 209  
 |||  
 Db 1 MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGGKELNRSRDIRIKYGNRKN SRLQFNKV 60  
 Qy 210 KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 269  
 |||  
 Db 61 KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 120  
 Qy 270 IEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQSVL 307  
 |||  
 Db 121 IEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQKHL 158

# RESULT 15

ABG71638

ID ABG71638 standard; protein; 407 AA.

XX

AC ABG71638;

XX

DT 14-JAN-2003 (first entry)

XX

DE Human membrane-bound splice variant of Don-1.

XX

KW Human; Don-1; epidermal growth factor; EGF; neuregulin;

KW glycoprotein ligand; cell proliferation; cell proliferative disorder;

KW carcinoma; adenocarcinoma cell; myeloma; cell differentiation;

KW cell survival; epithelial cell; wound healing; tumour formation; brain;





Search completed: August 17, 2004, 14:10:48  
Job time : 53.5478 secs

OM protein - protein search, using sw model

Run on: August 17, 2004, 14:09:05 ; Search time 16.2898 Seconds  
(without alignments)  
1045.842 Million cell updates/sec

Title: US-09-864-675-2  
Perfect score: 1749  
Sequence: 1 MRRDPAPGFSMLLFGVSLAC.....PGTGVSSSQWSTSPSTLDLN 330

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1722	98.5	330	2	US-08-525-864A-4	Sequence 4, Appli
2	1478	84.5	754	2	US-08-525-864A-2	Sequence 2, Appli
3	960	54.9	181	3	US-08-753-007A-4	Sequence 4, Appli
4	960	54.9	181	3	US-09-398-496-4	Sequence 4, Appli
5	881	50.4	469	3	US-08-753-007A-8	Sequence 8, Appli
6	881	50.4	469	3	US-09-398-496-8	Sequence 8, Appli
7	881	50.4	647	3	US-08-753-007A-32	Sequence 32, Appli
8	881	50.4	647	3	US-09-398-496-32	Sequence 32, Appli
9	842	48.1	407	3	US-08-753-007A-6	Sequence 6, Appli
10	842	48.1	407	3	US-09-398-496-6	Sequence 6, Appli
11	821	46.9	605	3	US-08-753-007A-2	Sequence 2, Appli

12	821	46.9	605	3	US-09-398-496-2	Sequence 2, Appli
13	748	42.8	139	3	US-08-753-007A-33	Sequence 33, Appl
14	748	42.8	139	3	US-09-398-496-33	Sequence 33, Appl
15	707	40.4	131	2	US-08-525-864A-6	Sequence 6, Appli
16	534.5	30.6	445	4	US-08-467-602-328	Sequence 328, App
17	534.5	30.6	479	4	US-08-467-602-370	Sequence 370, App
18	532.5	30.4	425	3	US-08-470-335-226	Sequence 226, App
19	532.5	30.4	425	4	US-08-467-602-320	Sequence 320, App
20	532.5	30.4	459	4	US-08-467-602-362	Sequence 362, App
21	532	30.4	414	3	US-08-470-339-188	Sequence 188, App
22	532	30.4	604	3	US-08-470-335-227	Sequence 227, App
23	532	30.4	604	4	US-08-467-602-318	Sequence 318, App
24	532	30.4	613	3	US-08-470-335-230	Sequence 230, App
25	532	30.4	613	4	US-08-467-602-329	Sequence 329, App
26	532	30.4	624	4	US-08-467-602-326	Sequence 326, App
27	532	30.4	633	4	US-08-467-602-335	Sequence 335, App
28	532	30.4	638	4	US-08-467-602-360	Sequence 360, App
29	532	30.4	647	4	US-08-467-602-371	Sequence 371, App
30	532	30.4	658	4	US-08-467-602-368	Sequence 368, App
31	532	30.4	667	4	US-08-467-602-377	Sequence 377, App
32	532	30.4	821	3	US-08-470-335-228	Sequence 228, App
33	532	30.4	821	4	US-08-467-602-319	Sequence 319, App
34	532	30.4	830	3	US-08-470-335-231	Sequence 231, App
35	532	30.4	830	4	US-08-467-602-330	Sequence 330, App
36	532	30.4	841	4	US-08-467-602-327	Sequence 327, App
37	532	30.4	850	4	US-08-467-602-336	Sequence 336, App
38	532	30.4	855	4	US-08-467-602-361	Sequence 361, App
39	532	30.4	864	4	US-08-467-602-372	Sequence 372, App
40	532	30.4	868	3	US-08-470-335-229	Sequence 229, App
41	532	30.4	868	4	US-08-467-602-317	Sequence 317, App
42	532	30.4	875	4	US-08-467-602-369	Sequence 369, App
43	532	30.4	877	3	US-08-470-335-232	Sequence 232, App
44	532	30.4	877	4	US-08-467-602-331	Sequence 331, App
45	532	30.4	884	4	US-08-467-602-378	Sequence 378, App

#### ALIGNMENTS

#### RESULT 1

US-08-525-864A-4

; Sequence 4, Application US/08525864A

; Patent No. 5912326

; GENERAL INFORMATION:

; APPLICANT: Chang, Han

; TITLE OF INVENTION: Cerebellum-derived Growth Factors, and Uses

; TITLE OF INVENTION: Related thereto

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: LAHIVE & COCKFIELD

; STREET: 28 State Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

```

;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  AscII (text)
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/525,864A
;      FILING DATE:  8-SEP-1995
;      CLASSIFICATION:  530
;      ATTORNEY/AGENT INFORMATION:
;      NAME:  Kara, Catherine J.
;      REGISTRATION NUMBER:  41,106
;      REFERENCE/DOCKET NUMBER:  HUI-017
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)742-4214
;      INFORMATION FOR SEQ ID NO:  4:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH:  330 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;      MOLECULE TYPE:  protein
US-08-525-864A-4

```

```

Query Match          98.5%;  Score 1722;  DB 2;  Length 330;
Best Local Similarity 97.9%;  Pred. No. 1.2e-147;
Matches 323;  Conservative 4;  Mismatches 3;  Indels 0;  Gaps 0;

```

```

QY      1 MRRDPAPGFSMLLFGVSLACYSPSLKSVDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MRRDPAPGFSMLLFGVSLACYSPSLKSVDQAYKAPVVVEGKVQGLAPAGGSSSNSTREP 60

QY     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAF 120
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAF 120

QY    121 PLDTNGKNLKKKEVGKILCTDCATRPKLKKMKMSQTGQVGEKQSLKCEAAAGNPQPSYRWF 180
      |:| |||:|||||||||||||||||||||||||:|||||||||||||||||
Db    121 PVDPNGKNIKKEVGKILCTDCATRPKLKKMKMSQTGEVGEKQSLKCEAAAGNPQPSYRWF 180

QY    181 DGKELNRSRDIRIKYNGRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLVNSVS 240
      |||||||||||||||||||||||||||||||||||||||||||:|||||
Db    181 DGKELNRSRDIRIKYNGRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLHVNSVS 240

QY    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP 300
      |||||||||||||||||||||||||||||||||||||||||||
Db    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP 300

QY    301 DPKQSVLWDTPGTGVSSSQWSTSPSTLDLN 330
      ||||||||||||||||||||||||
Db    301 DPKQSVLWDTPGTGVSSSQWSTSPSTLDLN 330

```

RESULT 2

US-08-525-864A-2

```

; Sequence 2, Application US/08525864A
; Patent No. 5912326
; GENERAL INFORMATION:
; APPLICANT:  Chang, Han

```

```

; TITLE OF INVENTION: Cerebellum-derived Growth Factors, and Uses
; TITLE OF INVENTION: Related thereto
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: AscII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,864A
; FILING DATE: 8-SEP-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Kara, Catherine J.
; REGISTRATION NUMBER: 41,106
; REFERENCE/DOCKET NUMBER: HUI-017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 754 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-525-864A-2

```

```

Query Match          84.5%; Score 1478; DB 2; Length 754;
Best Local Similarity 96.2%; Pred. No. 4.3e-125;
Matches 278; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

```

```

Qy      1 MRRDPAPGF SMLLF GVS LACYS PSLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
        |||
Db      1 MRRDPAPGF SMLLF GVS LACYS PSLKSVQDQAYKAPVVVEGKVQGLAPAGGSSSNSTREP 60

Qy     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 120
        |||
Db     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCAPLERNQRYIFFLEPTEQPLVFKTAFA 120

Qy    121 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180
        |:| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db    121 PVDPNGKNIKKEVGKILCTDCATRPKLKKMKSQTGEVGEKQSLKCEAAAGNPQPSYRWFK 180

Qy    181 DGKELNRSRDIRIKYGNRKN SRLQFNKV KVEDAGEYVCEAENILGKDTVGRRLVNSVS 240
        |||
Db    181 DGKELNRSRDIRIKYGNRKN SRLQFNKV KVEDAGEYVCEAENILGKDTVGRRLHVNSVS 240

Qy    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRC 289
        |||
Db    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDR 289

```

RESULT 3  
 US-08-753-007A-4  
 ; Sequence 4, Application US/08753007A  
 ; Patent No. 6074841  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gearing, David P.  
 ; APPLICANT: Busfield, Samantha J.  
 ; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES  
 ; TITLE OF INVENTION: AND USES THEREFOR  
 ; NUMBER OF SEQUENCES: 33  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Richardson P.C.  
 ; STREET: 225 Franklin Street  
 ; CITY: Boston  
 ; STATE: MA  
 ; COUNTRY: US  
 ; ZIP: 02110-2804  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSEQ Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/753,007A  
 ; FILING DATE: 19-NOV-1996  
 ; CLASSIFICATION: 536  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/699,591  
 ; FILING DATE: 19-AUG-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Fasse, J. Peter  
 ; REGISTRATION NUMBER: 32,983  
 ; REFERENCE/DOCKET NUMBER: 07334/022001  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-542-5070  
 ; TELEFAX: 617-542-8906  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 181 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: not relevant  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; FRAGMENT TYPE: internal  
 US-08-753-007A-4

Query Match 54.9%; Score 960; DB 3; Length 181;  
 Best Local Similarity 97.8%; Pred. No. 4.2e-79;  
 Matches 177; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 150 MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRSRDIRIKYGNRKN SRLQFNKV 209  
 |||||:|||||  
 Db 1 MKSQTGEVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRSRDIRIKYGNVRKN SRLQFNKV 60

Qy 210 KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 269  
 :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 61 RVEDAGEYVCEAENILGKDTVGRGLHVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 120  
 Qy 270 IEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQSVLWDTPGTGVSSSQWSTSPSTLDL 329  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 121 IEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQSVLWDTPGTGVSSSQWSTSPSTLDL 180  
 Qy 330 N 330  
 |  
 Db 181 N 181

RESULT 4

US-09-398-496-4

; Sequence 4, Application US/09398496  
 ; Patent No. 6133423  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gearing, David P.  
 ; APPLICANT: Busfield, Samantha J.  
 ; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES  
 ; TITLE OF INVENTION: AND USES THEREFOR  
 ; NUMBER OF SEQUENCES: 33  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Richardson P.C.  
 ; STREET: 225 Franklin Street  
 ; CITY: Boston  
 ; STATE: MA  
 ; COUNTRY: US  
 ; ZIP: 02110-2804  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSEQ Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/398,496  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/753,007  
 ; FILING DATE: 19-NOV-1996  
 ; APPLICATION NUMBER: 08/699,591  
 ; FILING DATE: 19-AUG-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Fasse, J. Peter  
 ; REGISTRATION NUMBER: 32,983  
 ; REFERENCE/DOCKET NUMBER: 07334/022001  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-542-5070  
 ; TELEFAX: 617-542-8906  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 181 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: not relevant





```

;     REGISTRATION NUMBER: 32,983
;     REFERENCE/DOCKET NUMBER: 07334/022001
;     TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 617-542-5070
;     TELEFAX: 617-542-8906
;     TELEX:
;     INFORMATION FOR SEQ ID NO: 8:
;     SEQUENCE CHARACTERISTICS:
;     LENGTH: 469 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: not relevant
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
;     FRAGMENT TYPE: internal
US-08-753-007A-8

```

```

Query Match          50.4%; Score 881; DB 3; Length 469;
Best Local Similarity 100.0%; Pred. No. 2.1e-71;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      142 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN 201
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31  ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN 90

Qy      202 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC 261
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91  SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC 150

Qy      262 VNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ 304
          ||||||||||||||||||||||||||||||||||||||||
Db      151 VNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ 193

```

# RESULT 6

US-09-398-496-8

```

; Sequence 8, Application US/09398496
; Patent No. 6133423
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/398,496

```

```

; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/753,007
; FILING DATE: 19-NOV-1996
; APPLICATION NUMBER: 08/699,591
; FILING DATE: 19-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07334/022001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 469 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-09-398-496-8

```

```

Query Match          50.4%; Score 881; DB 3; Length 469;
Best Local Similarity 100.0%; Pred. No. 2.1e-71;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      142 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRSRDIRIKYGNRKN 201
          |||
Db      31 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRSRDIRIKYGNRKN 90

Qy      202 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC 261
          |||
Db      91 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC 150

Qy      262 VNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ 304
          |||
Db      151 VNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ 193

```

# RESULT 7

US-08-753-007A-32

```

; Sequence 32, Application US/08753007A
; Patent No. 6074841
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA

```

```

; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/753,007A
; FILING DATE: 19-NOV-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/699,591
; FILING DATE: 19-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07334/022001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 647 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-753-007A-32

```

```

Query Match          50.4%; Score 881; DB 3; Length 647;
Best Local Similarity 100.0%; Pred. No. 3.3e-71;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      142 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRSRDIRIKYGNGRKN 201
          |||
Db      31 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRSRDIRIKYGNGRKN 90

Qy      202 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC 261
          |||
Db      91 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC 150

Qy      262 VNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ 304
          |||
Db      151 VNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ 193

```

```

RESULT 8
US-09-398-496-32
; Sequence 32, Application US/09398496
; Patent No. 6133423
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES

```

```

; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/398,496
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/753,007
; FILING DATE: 19-NOV-1996
; APPLICATION NUMBER: 08/699,591
; FILING DATE: 19-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07334/022001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 647 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-09-398-496-32

```

```

Query Match          50.4%;  Score 881;  DB 3;  Length 647;
Best Local Similarity 100.0%;  Pred. No. 3.3e-71;
Matches 163;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

Qy      142 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN 201
          |||
Db      31 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN 90

Qy      202 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC 261
          |||
Db      91 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC 150

Qy      262 VNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ 304
          |||
Db      151 VNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ 193

```

RESULT 9  
 US-08-753-007A-6  
 ; Sequence 6, Application US/08753007A  
 ; Patent No. 6074841  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gearing, David P.  
 ; APPLICANT: Busfield, Samantha J.  
 ; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES  
 ; TITLE OF INVENTION: AND USES THEREFOR  
 ; NUMBER OF SEQUENCES: 33  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Richardson P.C.  
 ; STREET: 225 Franklin Street  
 ; CITY: Boston  
 ; STATE: MA  
 ; COUNTRY: US  
 ; ZIP: 02110-2804  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSEQ Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/753,007A  
 ; FILING DATE: 19-NOV-1996  
 ; CLASSIFICATION: 536  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/699,591  
 ; FILING DATE: 19-AUG-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Fasse, J. Peter  
 ; REGISTRATION NUMBER: 32,983  
 ; REFERENCE/DOCKET NUMBER: 07334/022001  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-542-5070  
 ; TELEFAX: 617-542-8906  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 407 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: not relevant  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; FRAGMENT TYPE: internal  
 US-08-753-007A-6

Query Match 48.1%; Score 842; DB 3; Length 407;  
 Best Local Similarity 98.7%; Pred. No. 5.8e-68;  
 Matches 156; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 150 MKSQGTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNGRKNSRLQFNKV 209  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 MKSQGTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNGRKNSRLQFNKV 60  
 Qy 210 KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 269

```

Db          61 KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 120
Qy          270 IEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQSVL 307
Db          121 IEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQKHL 158

```

RESULT 10

US-09-398-496-6

```

; Sequence 6, Application US/09398496
; Patent No. 6133423
; GENERAL INFORMATION:
;   APPLICANT: Gearing, David P.
;   APPLICANT: Busfield, Samantha J.
;   TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
;   TITLE OF INVENTION: AND USES THEREFOR
;   NUMBER OF SEQUENCES: 33
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Fish & Richardson P.C.
;     STREET: 225 Franklin Street
;     CITY: Boston
;     STATE: MA
;     COUNTRY: US
;     ZIP: 02110-2804
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette
;     COMPUTER: IBM Compatible
;     OPERATING SYSTEM: DOS
;     SOFTWARE: FastSEQ Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/398,496
;     FILING DATE:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/753,007
;     FILING DATE: 19-NOV-1996
;     APPLICATION NUMBER: 08/699,591
;     FILING DATE: 19-AUG-1996
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Fasse, J. Peter
;     REGISTRATION NUMBER: 32,983
;     REFERENCE/DOCKET NUMBER: 07334/022001
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 617-542-5070
;     TELEFAX: 617-542-8906
;     TELEX:
;   INFORMATION FOR SEQ ID NO: 6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 407 amino acids
;       TYPE: amino acid
;       STRANDEDNESS: not relevant
;       TOPOLOGY: linear
;     MOLECULE TYPE: protein
;     FRAGMENT TYPE: internal

```

US-09-398-496-6

Query Match 48.1%; Score 842; DB 3; Length 407;  
Best Local Similarity 98.7%; Pred. No. 5.8e-68;  
Matches 156; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

[illegible]

RESULT 11

US-08-753-007A-2

; Sequence 2, Application US/08753007A

; Patent No. 6074841

## ; GENERAL INFORMATION:

; APPLICANT: Gearing, David P.

; APPLICANT: Busfield, Samantha J.

; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES

; TITLE OF INVENTION: AND USES THEREFOR

```
; NUMBER OF SEQUENCES: 33
```

## ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

```
; STREET: 225 Franklin Street
```

; CITY: Boston

; STATE: MA

; COUNTRY: US

; ZIP: 02110-2804

## ; COMPUTER READABLE FORM:

```
; MEDIUM TYPE: Diskette
```

```
;      COMPUTER:  IBM Compatible
```

```

;      OPERATING SYSTEM:  DOS

```

```
; SOFTWARE: FastSEQ Version 2.0
```

```
; CURRENT APPLICATION DATA:
```

; APPLICATION NUMBER: US/08/753,007A

FILING DATE: 19-NOV-1996

; CLASSIFICATION: 536

## ; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/699,591

; FILING DATE: 19-AUG-1996

## ; ATTORNEY/AGENT INFORMATION:

; NAME: Fasse, J. Peter

; REGISTRATION NUMBER: 32,983

REFERENCE/DOCKET NUMBER: 07334/022001

## ; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-542-5070

; TELEFAX: 617-542-8906

; TELEX:

## ; INFORMATION FOR SEO ID NO: 2:

## ; SEQUENCE CHARACTERISTICS:

; LENGTH: 605 amino acids



; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-08-753-007A-2

Query Match 46.9%; Score 821; DB 3; Length 605;  
Best Local Similarity 97.4%; Pred. No. 7.9e-66;  
Matches 151; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```
Qy      150 MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRSRDIRIKYGNRKN SRLQFNKV 209
          |||||:|||||
Db       1  MKSQTGEVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRSRDIRIKYGNVRKN SRLQFNKV 60

Qy      210 KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 269
          :|||||:|||||
Db       61 RVEDAGEYVCEAENILGKDTVGRGLHVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 120

Qy      270 IEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ 304
          |||||
Db      121 IEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ 155
```

RESULT 12

US-09-398-496-2

; Sequence 2, Application US/09398496  
; Patent No. 6133423  
; GENERAL INFORMATION:  
; APPLICANT: Gearing, David P.  
; APPLICANT: Busfield, Samantha J.  
; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES  
; TITLE OF INVENTION: AND USES THEREFOR  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/398,496  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/753,007  
; FILING DATE: 19-NOV-1996  
; APPLICATION NUMBER: 08/699,591  
; FILING DATE: 19-AUG-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fasse, J. Peter

```

;     REGISTRATION NUMBER: 32,983
;     REFERENCE/DOCKET NUMBER: 07334/022001
;     TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 617-542-5070
;     TELEFAX: 617-542-8906
;     TELEX:
;     INFORMATION FOR SEQ ID NO: 2:
;     SEQUENCE CHARACTERISTICS:
;     LENGTH: 605 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: not relevant
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
;     FRAGMENT TYPE: internal
US-09-398-496-2

```

```

Query Match          46.9%; Score 821; DB 3; Length 605;
Best Local Similarity 97.4%; Pred. No. 7.9e-66;
Matches 151; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      150 MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN SRLQFNKV 209
          |||||:|||||
Db      1   MKSQTGEVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNVRKN SRLQFNKV 60

Qy      210 KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 269
          :|||||:|||||
Db      61 RVEDAGEYVCEAENILGKDTVGRGLHVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 120

Qy      270 IEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ 304
          |||||
Db      121 IEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ 155

```

# RESULT 13

US-08-753-007A-33

```

; Sequence 33, Application US/08753007A
; Patent No. 6074841
; GENERAL INFORMATION:
;   APPLICANT: Gearing, David P.
;   APPLICANT: Busfield, Samantha J.
;   TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
;   TITLE OF INVENTION: AND USES THEREFOR
;   NUMBER OF SEQUENCES: 33
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson P.C.
;   STREET: 225 Franklin Street
;   CITY: Boston
;   STATE: MA
;   COUNTRY: US
;   ZIP: 02110-2804
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/753,007A

```

```

; FILING DATE: 19-NOV-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/699,591
; FILING DATE: 19-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07334/022001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX:
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 139 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-753-007A-33

```

```

Query Match          42.8%; Score 748; DB 3; Length 139;
Best Local Similarity 98.6%; Pred. No. 4.3e-60;
Matches 137; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      192 RIKYGNRGRKNSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHAR 251
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      1  RIKYGNRGRKNSRLQFNKVRVEDAGEYVCEAENILGKDTVGRGLHVNSVSTTLSSWSGHAR 60

Qy      252 KCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQSVLWDTP 311
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      61 KCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQSVLWDTP 120

Qy      312 GTGVSSSQWSTSPSTLDLN 330
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      121 GTGVSSSQWSTSPSTLDLN 139

```

#### RESULT 14

US-09-398-496-33

```

; Sequence 33, Application US/09398496
; Patent No. 6133423
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:

```



```

;   ADDRESSEE:  LAHIVE & COCKFIELD
;   STREET:  28 State Street
;   CITY:  Boston
;   STATE:  Massachusetts
;   COUNTRY:  USA
;   ZIP:  02109
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE:  Floppy disk
;   COMPUTER:  IBM PC compatible
;   OPERATING SYSTEM:  PC-DOS/MS-DOS
;   SOFTWARE:  AscII (text)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/08/525,864A
;   FILING DATE:  8-SEP-1995
;   CLASSIFICATION:  530
;   ATTORNEY/AGENT INFORMATION:
;   NAME:  Kara, Catherine J.
;   REGISTRATION NUMBER:  41,106
;   REFERENCE/DOCKET NUMBER:  HUI-017
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  (617)227-7400
;   TELEFAX:  (617)742-4214
;   INFORMATION FOR SEQ ID NO:  6:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH:  131 amino acids
;   TYPE:  amino acid
;   TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
US-08-525-864A-6

```

```

Query Match          40.4%;  Score 707;  DB 2;  Length 131;
Best Local Similarity 99.2%;  Pred. No. 2e-56;
Matches 130; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      200 KNSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKS 259
          |||||||||||||||||||||||||||||||||||||||:||||||||||||||||||
Db      1   KNSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLHVNSVSTTLSSWSGHARKCNETAKS 60

Qy      260 YCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQSVLWDTPGTGVSQQ 319
          |||||||||||||||||||||||||||||||||||||||
Db      61 YCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQSVLWDTPGTGVSQQ 120

Qy      320 WSTSPSTLDLN 330
          |||||||||
Db      121 WSTSPSTLDLN 131

```

```

Search completed: August 17, 2004, 14:14:01
Job time : 17.2898 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 17, 2004, 14:06:50 ; Search time 15.2389 Seconds  
(without alignments)  
2083.044 Million cell updates/sec

Title: US-09-864-675-2  
Perfect score: 1749  
Sequence: 1 MRRDPAPGFSSMLLFGVSLAC.....PGTGVSSSQWSTSPSTLDLN 330

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	1610	92.1	850	2	JC5700	ErbB kinase activa
2	1576	90.1	868	2	JC5701	ErbB kinase activa
3	1575	90.1	860	2	JC5702	ErbB kinase activa
4	523	29.9	422	2	S32357	glial growth facto
5	293.5	16.8	462	2	I38404	neu differentiatio
6	292.5	16.7	640	2	A43273	heregulin precurso
7	286	16.4	639	2	I61719	neu differentiatio
8	285.5	16.3	241	2	D43273	heregulin precurso
9	285.5	16.3	637	2	C43273	heregulin precurso
10	285.5	16.3	645	2	B43273	heregulin, splice
11	284.5	16.3	602	2	A45769	acetylcholine rece
12	283	16.2	662	2	I61722	neu differentiatio
13	278	15.9	230	2	A56210	neu differentiatio

14	278	15.9	636	2	I61718	neu differentiatio
15	271	15.5	241	2	S32359	glial growth facto
16	208.5	11.9	125	2	S62676	heregulin isoform
17	190.5	10.9	175	2	I38408	neu differentiatio
18	164.5	9.4	5175	2	T20992	hypothetical prote
19	164.5	9.4	5198	2	T43290	hemicentin precurs
20	151	8.6	1323	2	PN0568	connectin 3B - chi
21	150.5	8.6	6658	2	T13931	projectin - fruit
22	146	8.3	125	2	I38405	neu differentiatio
23	144.5	8.3	7962	2	I38346	elastic titin - hu
24	144	8.2	338	2	JC4776	limbic-system-asso
25	143.5	8.2	1612	2	T30805	duttl protein - mo
26	142.5	8.1	919	2	T32541	unc-5 protein - Ca
27	142.5	8.1	947	1	B44294	unc-5 protein, lon
28	142	8.1	2783	2	T34416	hypothetical prote
29	140.5	8.0	1651	2	T14160	transmembrane rece
30	140	8.0	761	1	IJHUNG	neural cell adhesi
31	139.5	8.0	6642	2	T29757	protein UNC-89 - C
32	139	7.9	1896	2	T08851	Down syndrome cell
33	135.5	7.7	1356	2	JC1402	protein-tyrosine k
34	135	7.7	4391	2	A38096	perlecan precursor
35	135	7.7	26926	1	I38344	titin, cardiac mus
36	133.5	7.6	853	1	IJBONC	neural cell adhesi
37	133.5	7.6	1367	2	A41228	protein-tyrosine k
38	132.5	7.6	4162	2	T42633	connectin/titin -
39	131	7.5	296	2	A56943	sensory/motor neur
40	131	7.5	338	2	JC1238	opioid-binding pro
41	131	7.5	345	2	JC1239	opioid-binding pro
42	130.5	7.5	940	2	A40985	projectin - fruit
43	130.5	7.5	1040	2	A34695	axonal glycoprotei
44	129.5	7.4	1018	2	A54744	contactin 1 precur
45	129.5	7.4	1091	1	IJCHNL	neural cell adhesi

#### ALIGNMENTS

##### RESULT 1

JC5700

ErbB kinase activator alpha, brain and thymus - human

C;Species: Homo sapiens (man)

C;Date: 25-Nov-1997 #sequence\_revision 25-Nov-1997 #text\_change 08-Sep-2002

C;Accession: JC5700

R;Higashiyama, S.; Horikawa, M.; Yamada, K.; Ichino, N.; Nakano, N.; Nakagawa, T.; Miyagawa, J.; Matsushita, N.; Nagatsu, T.; Taniguchi, N.; Ishiguro, H. J. Biochem. 122, 675-680, 1997

A;Title: A novel brain-derived member of the epidermal growth factor family that interacts with ErbB3 and ErbB4.

A;Reference number: JC5700; MUID:98006324; PMID:9348101

A;Accession: JC5700

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-850 <HIG>

A;Cross-references: DDBJ:AB005060; NID:g2626738; PIDN:BAA23417.1; PID:g2626739

A;Experimental source: SK-NSH cell

C;Comment: This protein is a member of the epidermal growth factor family. It is functionally similar to neurogulin in terms of directly activating ErbB4,

transactivating ErbB1, B2 and B3, and stimulating the differentiation of MDA-MB-453 cells.

C;Superfamily: human ErbB kinase activator alpha, brain and thymus; EGF homology; immunoglobulin homology

C;Keywords: glycoprotein

F;258-311/Domain: Ig-like #status predicted <IGL>

F;345-381/Domain: EGF homology <EGF>

F;346-381/Domain: EGF-like #status predicted <EGF2>

F;147,278,451/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 92.1%; Score 1610; DB 2; Length 850;

Best Local Similarity 100.0%; Pred. No. 1.1e-114;

Matches 304; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
      |||
Db     93 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 152

Qy     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAF 120
      |||
Db    153 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAF 212

Qy    121 PLDTNGKNLKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWF 180
      |||
Db    213 PLDTNGKNLKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWF 272

Qy    181 DGKELNRSRDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGRLYVNSVS 240
      |||
Db    273 DGKELNRSRDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGRLYVNSVS 332

Qy    241 TTLSSWSGHARKCNETAKSYCVNGGVVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP 300
      |||
Db    333 TTLSSWSGHARKCNETAKSYCVNGGVVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP 392

Qy    301 DPKQ 304
      |||
Db    393 DPKQ 396
```

## RESULT 2

JC5701

ErbB kinase activator alpha1, brain and thymus - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 25-Nov-1997 #sequence\_revision 25-Nov-1997 #text\_change 08-Sep-2002

C;Accession: JC5701; PC4411

R;Higashiyama, S.; Horikawa, M.; Yamada, K.; Ichino, N.; Nakano, N.; Nakagawa, T.; Miyagawa, J.; Matsushita, N.; Nagatsu, T.; Taniguchi, N.; Ishiguro, H. J. Biochem. 122, 675-680, 1997

A;Title: A novel brain-derived member of the epidermal growth factor family that interacts with ErbB3 and ErbB4.

A;Reference number: JC5700; MUID:98006324; PMID:9348101

A;Accession: JC5701

A;Molecule type: mRNA

A;Residues: 1-868 <HIG>

A;Cross-references: DDBJ:D89995; NID:g2605629; PIDN:BAA23344.1; PID:g2605630

A;Accession: PC4411

A;Molecule type: protein



F;361-397/Domain: EGF homology <EGF>

Query Match 90.1%; Score 1576; DB 2; Length 868;

Best Local Similarity 96.7%; Pred. No. 4.4e-112;

Matches 297; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy	1	MRRDPAPGFSMLLFGVSLACYSPSLKSVDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP	60
Db	109	MRRDPAPGSSMLLFGVSLACYSPSLKSVDQAYKAPVVVEGKVQGLAPAGGSSSNSTREP	168
Qy	61	PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAF	120
Db	169	PASGRVALVKVLDKWPLRSGGLQREQVISVGSCAPLERNQRYIFFLEPTEQPLVFKTAF	228
Qy	121	PLDTNGKNLKKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWF	180
Db	229	PVDPNGKNIKKEVGKILCTDCATRPKLKKMKSQTGVEGKQSLKCEAAAGNPQPSYRWF	288
Qy	181	DGKELNRSRDIRIKYGNRKNLSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLVNSVS	240
Db	289	DGKELNRSRDIRIKYGNRKNLSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLHVNVS	348
Qy	241	TTLSSWSGHARKCNETAKSYCVNGGVCYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP	300
Db	349	TTLSSWSGHARKCNETAKSYCVNGGVCYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP	408
Qy	301	DPKQSVL 307	
Db	409	DPKQKHL 415	

### RESULT 3

JC5702

ErbB kinase activator alpha2a, brain and thymus - rat

C;Species: Rattus norvegicus (Norway rat)

C:Date: 25-Nov-1997 #sequence revision 25-Nov-1997 #text\_change 08-Sep-2002

C;Accession: JC5702; PC4417

R; Higashiyama, S.; Horikawa, M.; Yamada, K.; Ichino, N.; Nakano, N.; Nakagawa, T.; Miyagawa, J.; Matsushita, N.; Nagatsu, T.; Taniguchi, N.; Ishiguro, H. J. Biochem. 122, 675-680, 1997

A;Title: A novel brain-derived member of the epidermal growth factor family that interacts with ErbB3 and ErbB4.

A:Reference number: JC5700; MUID:98006324; PMID:9348101

A;Accession: JC5702

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-860 &lt;HIG&gt;

A;Cross-references: DDBJ:D89996; NID:q2605631; PIDN:BAA23345.1; PID:g2605632

A; Experimental source: PC-12 cell

A;Accession: PC4417  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 'F',212-213,223-860 <HI2>  
A;Cross-references: DDBJ:AB001576; NID:g2605478; PIDN:BAA23348.1; PID:g2605479  
A;Experimental source: PC-12 cell  
C;Comment: This protein is a member of the epidermal growth factor family. It is functionally similar to neurogulin in terms of directly activating ErbB4, transactivating ErbB1, B2 and B3, and stimulating the differentiation of MDA-MB-453 cells.  
C;Superfamily: human ErbB kinase activator alpha, brain and thymus; EGF homology; immunoglobulin homology  
C;Keywords: glycoprotein  
F;274-327/Domain: Ig-like #status predicted <IGL>  
F;361-397/Domain: EGF homology <EGF>  
F;422-444/Domain: hydrophobic #status predicted <HYD>  
F;163,294,467/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 90.1%; Score 1575; DB 2; Length 860;  
Best Local Similarity 97.4%; Pred. No. 5.2e-112;  
Matches 296; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```

Qy      1 MRRDPAPGFSMLLFGVSLACYSPSLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      109 MRRDPAPGSSMLLFGVSLACYSPSLKSVQDQAYKAPVVVEGKVQGLAPAGGSSSNSTREP 168

Qy      61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAF 120
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      169 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCAPLERNQRYIFFLEPTEQPLVFKTAF 228

Qy      121 PLDTNGKNLKKKEVGKILCTDCATRPKLKKMKSQGTGQVGEKQSLKCEAAAGNPQPSYRWFK 180
          |:|  ||||:|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      229 PVDPNGKNIKKEVGKILCTDCATRPKLKKMKSQGTGEVGEKQSLKCEAAAGNPQPSYRWFK 288

Qy      181 DGKELNRSRDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRRLYVNSVS 240
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      289 DGKELNRSRDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRRLHVNSVS 348

Qy      241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP 300
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      349 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP 408

Qy      301 DPKQ 304
          ||||
Db      409 DPKQ 412

```

#### RESULT 4

S32357

glial growth factor - human

C;Species: Homo sapiens (man)

C;Date: 02-Dec-1993 #sequence\_revision 10-Nov-1995 #text\_change 08-Sep-2002

C;Accession: S32357

R;Marchionni, M.A.; Goodearl, A.D.J.; Chen, M.S.; Bermingham-McDonogh, O.; Kirk, C.; Hendricks, M.; Danehy, F.; Misumi, D.; Sudhalter, J.; Kobayashi, K.; Wroblewski, D.; Lynch, C.; Baldassare, M.; Hiles, I.; Davis, J.B.; Hsuan, J.J.;

Totty, N.F.; Otsu, M.; McBurney, R.N.; Waterfield, M.D.; Stroobant, P.; Gwynne, D.  
 Nature 362, 312-318, 1993  
 A;Title: Glial growth factors are alternatively spliced erbB2 ligands expressed in the nervous system.  
 A;Reference number: S32357; MUID:93205115; PMID:8096067  
 A;Accession: S32357  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-422 <MAR>  
 A;Cross-references: GB:L12260; NID:g292047; PIDN:AAB59622.1; PID:g292048  
 C;Superfamily: human heregulin; EGF homology; immunoglobulin homology  
 F;363-402/Domain: EGF homology <EGF>

Query Match 29.9%; Score 523; DB 2; Length 422;  
 Best Local Similarity 35.5%; Pred. No. 2.9e-32;  
 Matches 124; Conservative 59; Mismatches 88; Indels 78; Gaps 13;

```

Qy      15 GVSLACYS--PSLKSVQDQAYKAPVVVEGKV-----QGLV-----PAGGSSS--NSTRE 59
      | | : ||| || : ||| : | : | || : ||| || : || : ||
Db      58 GASV-CYSSPPSVGSGVQELAQRAAVVIEGKVHPQRRQQGALDRKAAAAAGEAGAWGGDRE 116

Qy      60 PPASGRVA-----LVKVLDKWPLRSGGLQ 83
      ||| : | ||| || : ||| :
Db     117 PPAAGPRALGPPAEELLAANGTVPSWPTAPVPSAGEPGEEAPYLVKVHQVWAVKAGGLK 176

Qy      84 REQVISV-----GSCVPLERNQRYIFFLEP-----TEQPLVFKTAFAPLDTNGKN 128
      :: ::| || | : : |||| : || : | | : : | || : | : |
Db     177 KDSLLTVRLGTWGHPAFPSCGRCLKEDSRYIFFMEPDANSTSRAPAAFRASFPPLET-GRN 235

Qy     129 LKKEVGKILCTDCATRPKLKKMKSQGTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRS 188
      ||||| ::|| || | : || : ||| || | | : || : : : ||| : |||
Db     236 LKKEVSRVLCKRCALPPQLKEMKSQESAAGSKLVLRCESTSSEYSSLRFKWFKNGNELNRK 295

Qy     189 ---RDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSS 245
      ::| : | : | | : | : ||| : | : | | : : : | : |
Db     296 NKPQNIKIQKKPGK--SELRINKASLADSGEYMCKVISKLGNDSASANITIVESNATSTS 353

Qy     246 WSG--HARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPNGFFGQRC 289
      : | | || | | : ||||| | : : : | |||| | | ||
Db     354 TTGTSHLVKCAEKEKTFCVNGGECFMVKDLNPSRYLCKCPNEFTGDRC 402

```

# RESULT 5

I38404

neu differentiation factor - human

C;Species: Homo sapiens (man)

C;Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 08-Sep-2002

C;Accession: I38404

R;Wen, D.; Suggs, S.V.; Karunakaran, D.; Liu, N.; Cupples, R.L.; Luo, Y.; Janssen, A.M.; Ben-Baruch, N.; Trollinger, D.B.; Jacobsen, V.L.; Meng, S. Mol. Cell. Biol. 14, 1909-1919, 1994

A;Title: Structural and functional aspects of the multiplicity of Neu differentiation factors.

A;Reference number: A56210; MUID:94158863; PMID:7509448

A;Accession: I38404

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA  
A;Residues: 1-462 <RES>  
A;Cross-references: EMBL:U02326; NID:g408402; PIDN:AAA19951.1; PID:g408403  
C;Superfamily: human heregulin; EGF homology; immunoglobulin homology

Query Match 16.8%; Score 293.5; DB 2; Length 462;  
Best Local Similarity 32.1%; Pred. No. 9.6e-15;  
Matches 71; Conservative 37; Mismatches 62; Indels 51; Gaps 7;

```
Qy      126 GKNLKKEVGKILCTDCAT-----RPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRW 178
      || ||| | : | |::||| | | |::| ::|
Db      11 GKGGKKKRGSGKKPESAAGSQSPALPPRLKEMKSQESAAGSKLVLRCESSSEYSSLRFKW 70

Qy      179 FKDGKELNRS---RDIRIKYGNRKNRLQFNKVKVEDAGEYVCEAENILGKDTVGRGL- 234
      ||:| ||| | ::||: | : | | | : |::|:| : || | :
Db      71 FKNGNELNRKNKPQNIKIQKKPGK--SELRINKASLADSGEYMCKVISKLGNDSSASANIT 128

Qy      235 -----YV-----NSVSTTLSSWSG--HARKCNETAKS 259
      || |::|:|:| | || | |:
Db      129 IVESNEIITGMPASTEGAYVSSESPIRISVSTEGANTSSSTSTSTGTSHLVKCAEKEKT 188

Qy      260 YCVNGGVCYYIEGINQLS---CKCPNGFFGQRCLEKLPLRL 297
      :||| | |::: | || | | | |:::
Db      189 FCVNGGECFMVKDLSNPSRYLCKCQPGFTGARCTENVPMKV 229
```

#### RESULT 6

A43273

heregulin precursor, splice form alpha - human

N;Alternate names: breast cancer cell differentiation factor p45; Neu differentiation factor

C;Species: Homo sapiens (man)

C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 08-Sep-2002

C;Accession: A43273; A48498; A38155

R;Holmes, W.E.; Sliwkowski, M.X.; Akita, R.W.; Henzel, W.J.; Lee, J.; Park, J.W.; Yansura, D.; Abadi, N.; Raab, H.; Lewis, G.D.; Shepard, H.M.; Kuang, W.J.; Wood, W.I.; Goeddel, D.V.; Vandlen, R.L.  
Science 256, 1205-1210, 1992

A;Title: Identification of heregulin, a specific activator of p185(erbB2).

A;Reference number: A43273; MUID:92271253; PMID:1350381

A;Accession: A43273

A;Status: nucleic acid sequence not shown; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-640 <HOL>

A;Experimental source: breast tumor cell line, MDA-MB-231, ATCC HTB 26

A;Note: sequence extracted from NCBI backbone (NCBIP:103250)

R;Culouscou, J.M.; Plowman, G.D.; Carlton, G.W.; Green, J.M.; Shoyab, M.  
J. Biol. Chem. 268, 18407-18410, 1993

A;Title: Characterization of a breast cancer cell differentiation factor that specifically activates the HER4/p180(erbB4) receptor.

A;Reference number: A48498; MUID:93366731; PMID:7689552

A;Accession: A48498

A;Molecule type: protein

A;Residues: 20-21, 'X', 23-24, 'XX', 27-28 <CUL>

R;Peles, E.; Bacus, S.S.; Koski, R.A.; Lu, H.S.; Wen, D.; Ogden, S.G.; Levy, R.B.; Yarden, Y.

Cell 69, 205-216, 1992

A;Title: Isolation of the neu/HER-2 stimulatory ligand: a 44 kd glycoprotein that induces differentiation of mammary tumor cells.

A;Reference number: A38155; MUID:92208945; PMID:1348215

A;Accession: A38155

A;Molecule type: protein

A;Residues: 'X',15-16,'X',18-20,'RG',23-24,'GP',27,'E',29,'XP',32-36 <PEL>

A;Note: sequence extracted from NCBI backbone (NCBIP:91347)

C;Genetics:

A;Gene: GDB:HGL

A;Cross-references: GDB:132656; OMIM:142445

A;Map position: 8p22-8p11

C;Superfamily: human heregulin; EGF homology; immunoglobulin homology

C;Keywords: alternative splicing; glycoprotein

F;182-221/Domain: EGF homology <EGF>

Query Match 16.7%; Score 292.5; DB 2; Length 640;

Best Local Similarity 32.1%; Pred. No. 1.7e-14;

Matches 71; Conservative 37; Mismatches 62; Indels 51; Gaps 7;

```
Qy      126 GKNLKKEVGKILCTDCAT-----RPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRW 178
      || ||| | : | |::| || | | :|| :: :|
Db      11 GKGGKKKRGSGKKPESAAGSQSPALPPQLKEMKSQESAAGSKLVLCETSSEYSSLRFKW 70

Qy      179 FKDGKELNRS---RDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGL- 234
      ||:| |||| ::||:| |: | |: || : |:| |:| : : || |: :
Db      71 FKNGNELNRKNKPQNIKIQQKPGK--SELRINKASLADSGEYMCKVISKLGNDSSASANIT 128

Qy      235 -----YV-----NSVSTTLSSWSG--HARKCNETAKS 259
      || |::| :| :| | || | |:
Db      129 IVESNEIITGMPASTEGAYVSSSPIRISVSTEGANTSSSTSTTTGTSHLVKCAEKEKT 188

Qy      260 YCVNGGVCYYIEGINQLS---CKCPNGFFGQRCLEKLPLRL 297
      :||| || : : : | || | || | :|::
Db      189 FCVNGGECFMVKDLSNPSRYLCKCQPGFTGARCTENVPMKV 229
```

#### RESULT 7

I61719

neu differentiation factor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 08-Sep-2002

C;Accession: I61719; I61723; I61716; I61717; I61724; A38220

R;Wen, D.; Suggs, S.V.; Karunakaran, D.; Liu, N.; Cupples, R.L.; Luo, Y.;

Janssen, A.M.; Ben-Baruch, N.; Trollinger, D.B.; Jacobsen, V.L.; Meng, S.

Mol. Cell. Biol. 14, 1909-1919, 1994

A;Title: Structural and functional aspects of the multiplicity of Neu differentiation factors.

A;Reference number: A56210; MUID:94158863; PMID:7509448

A;Accession: I61719

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-639 <RES>

A;Cross-references: EMBL:U02319; NID:g408388; PIDN:AAA19944.1; PID:g408389

A;Accession: I61723

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-639 <RE2>  
 A;Cross-references: EMBL:U02323; NID:g408396; PIDN:AAA19948.1; PID:g408397  
 A;Accession: I61716  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-422, 'H', 'NL', 637-638, 'ELRRNKAYRSKCMQIQLSATHLRPSSITHLGFI' <RE3>  
 A;Cross-references: EMBL:U02316; NID:g408382; PIDN:AAA19941.1; PID:g408383  
 A;Accession: I61717  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-422, 'H', 'NL', 637-638, 'ELRRNKAYRSKCMQIQLSATHLRPSSITHLGFI' <RE4>  
 A;Cross-references: EMBL:U02317; NID:g408384; PIDN:AAA19942.1; PID:g408385  
 A;Accession: I61724  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-422 <RE5>  
 A;Cross-references: EMBL:U02324; NID:g408398; PIDN:AAA19949.1; PID:g408399  
 R;Wen, D.; Peles, E.; Cupples, R.; Suggs, S.V.; Bacus, S.S.; Luo, Y.; Trail, G.;  
 Hu, S.; Silbiger, S.M.; Levy, R.B.; Koski, R.A.; Lu, H.S.; Yarden, Y.  
 Cell 69, 559-572, 1992  
 A;Title: Neu differentiation factor: a transmembrane glycoprotein containing an  
 EGF domain and an immunoglobulin homology unit.  
 A;Reference number: A38220; MUID:92257596; PMID:1349853  
 A;Accession: A38220  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-422 <WEN>  
 A;Note: sequence extracted from NCBI backbone (NCBIN:101767, NCBIP:101768)  
 C;Superfamily: human heregulin; EGF homology; immunoglobulin homology

Query Match 16.4%; Score 286; DB 2; Length 639;  
 Best Local Similarity 32.8%; Pred. No. 5.3e-14;  
 Matches 65; Conservative 35; Mismatches 54; Indels 44; Gaps 6;

```

Qy      142 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRS---RDIRIKYNGNG 198
      |  |::|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      34 ALPRLKEMKSQESAAGSKLVLCETSSSEYSSLRFKWFKN GNELNRKNKPENIKIQKKPG 93

Qy      199 RKNSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGL-----YV--- 236
      :  | | : | | : | | | | | | | | | | | | | | | | | | | | | |
Db      94 K--SELRINKASLADSGEYMCKVISKLGNDSASANITIVESNEFITGMPASTETAYVSSE 151

Qy      237 -----NSVSTTLSSWSG--HARKCNETAKSYCVNGGVCYYIEGINQLS---CK 279
      | : | : | : | : | | | | | | | | | | | | | | | | | | | | | |
Db      152 SPIRISVSTEGANTSSSTSTTGTSTHSLIKCAEKEKTFVNGGECFTVKDLSPRYLCK 211

Qy      280 CPNGFFGQRCLEKLPLRL 297
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      212 CQPGFTGARCTENVPMKV 229
  
```

RESULT 8

D43273

heregulin precursor, splice form beta-3 - human

N;Alternate names: glial growth factor HRG-beta-3; neuregulin

C;Species: Homo sapiens (man)

C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 08-Sep-2002

C;Accession: D43273; S32358  
R;Holmes, W.E.; Sliwkowski, M.X.; Akita, R.W.; Henzel, W.J.; Lee, J.; Park, J.W.; Yansura, D.; Abadi, N.; Raab, H.; Lewis, G.D.; Shepard, H.M.; Kuang, W.J.; Wood, W.I.; Goeddel, D.V.; Vandlen, R.L.  
Science 256, 1205-1210, 1992  
A;Title: Identification of heregulin, a specific activator of p185(erbB2).  
A;Reference number: A43273; MUID:92271253; PMID:1350381  
A;Accession: D43273  
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-241 <HOL>  
R;Marchionni, M.A.; Goodearl, A.D.J.; Chen, M.S.; Bermingham-McDonogh, O.; Kirk, C.; Hendricks, M.; Danehy, F.; Misumi, D.; Sudhalter, J.; Kobayashi, K.; Wroblewski, D.; Lynch, C.; Baldassare, M.; Hiles, I.; Davis, J.B.; Hsuan, J.J.; Totty, N.F.; Otsu, M.; McBurney, R.N.; Waterfield, M.D.; Stroobant, P.; Gwynne, D.  
Nature 362, 312-318, 1993  
A;Title: Glial growth factors are alternatively spliced erbB2 ligands expressed in the nervous system.  
A;Reference number: S32357; MUID:93205115; PMID:8096067  
A;Accession: S32358  
A;Molecule type: mRNA  
A;Residues: 1-241 <MAR>  
A;Cross-references: GB:L12261; NID:g292049; PIDN:AAB59358.1; PID:g292050  
C;Genetics:  
A;Gene: GDB:HGL; GGF  
A;Cross-references: GDB:132656; OMIM:142445  
A;Map position: 8p22-8p11  
C;Superfamily: human heregulin; EGF homology; immunoglobulin homology  
C;Keywords: alternative splicing  
F;182-221/Domain: EGF homology <EGF>

Query Match 16.3%; Score 285.5; DB 2; Length 241;  
Best Local Similarity 32.9%; Pred. No. 1.8e-14;  
Matches 70; Conservative 33; Mismatches 59; Indels 51; Gaps 7;

```

Qy      126 GKNLKKEVKGILCTDCAT-----RPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRW 178
      || ||| | : | |::||| | | |::| :: |
Db      11 GKGKKKERGSGKPPESAAGSQSPALPPRLKEMKSQESAAGSKLVLRCESSSEYSSLRFKW 70

Qy      179 FKDGKELNRS---RDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGL- 234
      ||::| ||| | ::::| | : | | | : |::|::| : || | :
Db      71 FKNGNELNRKNKPQNIKIQQKPGK--SELRINKASLADSGEYMCKVISKLGNDSASANIT 128

Qy      235 -----YV-----NSVSTTLSSWSG--HARKCNETAKS 259
      || |::| :| :| | || | |
Db      129 IVESNEIITGMPASTEGAYVSSESPIRISVSTEGANTSSSTSTTGTSHLVKCAEKEKT 188

Qy      260 YCVNGGVCYIEGINQLS---CKCPNGFFGQRC 289
      :||| | |::: | |||| | | |
Db      189 FCVNGGECFMVKDLSNPSRYLCKCPNEFTGDRC 221

```

RESULT 9

C43273

heregulin precursor, splice form beta-2 - human

C;Species: Homo sapiens (man)  
 C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 08-Sep-2002  
 C;Accession: C43273; I38407  
 R;Holmes, W.E.; Sliwkowski, M.X.; Akita, R.W.; Henzel, W.J.; Lee, J.; Park, J.W.; Yansura, D.; Abadi, N.; Raab, H.; Lewis, G.D.; Shepard, H.M.; Kuang, W.J.; Wood, W.I.; Goeddel, D.V.; Vandlen, R.L.  
 Science 256, 1205-1210, 1992  
 A;Title: Identification of heregulin, a specific activator of p185(erbB2).  
 A;Reference number: A43273; MUID:92271253; PMID:1350381  
 A;Accession: C43273  
 A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation  
 A;Molecule type: mRNA  
 A;Residues: 1-637 <HOL>  
 R;Wen, D.; Suggs, S.V.; Karunagaran, D.; Liu, N.; Cupples, R.L.; Luo, Y.; Janssen, A.M.; Ben-Baruch, N.; Trollinger, D.B.; Jacobsen, V.L.; Meng, S.  
 Mol. Cell. Biol. 14, 1909-1919, 1994  
 A;Title: Structural and functional aspects of the multiplicity of Neu differentiation factors.  
 A;Reference number: A56210; MUID:94158863; PMID:7509448  
 A;Accession: I38407  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: mRNA  
 A;Residues: 119-406 <RES>  
 A;Cross-references: EMBL:U02329; NID:g408408; PIDN:AAA19954.1; PID:g408409  
 C;Genetics:  
 A;Gene: GDB:HGL  
 A;Cross-references: GDB:132656; OMIM:142445  
 A;Map position: 8p22-8p11  
 C;Superfamily: human heregulin; EGF homology; immunoglobulin homology  
 C;Keywords: alternative splicing  
 F;182-221/Domain: EGF homology <EGF>

Query Match 16.3%; Score 285.5; DB 2; Length 637;  
 Best Local Similarity 32.9%; Pred. No. 5.7e-14;  
 Matches 70; Conservative 33; Mismatches 59; Indels 51; Gaps 7;

```

QY      126 GKNLKKEVGKILCTDCAT-----RPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRW 178
          || ||| | : | |::||| | | |::| :: ::|
Db      11 GKGGKKKERGSGKKPESAAGSQSPALPPRLKEMKSQESAAGSKLVLRCESSSEYSSLRFKW 70

QY      179 FKDGKELNRS---RDIRIKYGNRKNRLQFNKVKVEDAGEYVCEAENILGKDTVGRGL- 234
          ||:| |||| ::||: | : | | : |::||: : || | :
Db      71 FKNGNELNRKNKPQNIKIQKKPGK--SELRINKASLADSGEYMCKVISKLGNDSSASANIT 128

QY      235 -----YV-----NSVSTTLSSWSG--HARKCNETAKS 259
          || |::| :| :| | || | |
Db      129 IVESNEIITGMPASTEGAYVSSESPIRISVSTEGANTSSSTSTSTGTSHLVKCAEKEKT 188

QY      260 YCVNGGVCCYYIEGINQLS---CKCPNGFFGQRC 289
          :||| | : :: : | |||| | | |
Db      189 FCVNGGECFMVKDLNPSRYLCKCPNEFTGDRC 221
  
```

RESULT 10  
 B43273  
 heregulin, splice form beta 1 - human



C;Species: Homo sapiens (man)  
 C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 08-Sep-2002  
 C;Accession: B43273; I38406  
 R;Holmes, W.E.; Sliwkowski, M.X.; Akita, R.W.; Henzel, W.J.; Lee, J.; Park, J.W.; Yansura, D.; Abadi, N.; Raab, H.; Lewis, G.D.; Shepard, H.M.; Kuang, W.J.; Wood, W.I.; Goeddel, D.V.; Vandlen, R.L.  
 Science 256, 1205-1210, 1992  
 A;Title: Identification of heregulin, a specific activator of p185(erbB2).  
 A;Reference number: A43273; MUID:92271253; PMID:1350381  
 A;Accession: B43273  
 A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation  
 A;Molecule type: mRNA  
 A;Residues: 1-645 <HOL>  
 R;Wen, D.; Suggs, S.V.; Karunakaran, D.; Liu, N.; Cupples, R.L.; Luo, Y.; Janssen, A.M.; Ben-Baruch, N.; Trollinger, D.B.; Jacobsen, V.L.; Meng, S.  
 Mol. Cell. Biol. 14, 1909-1919, 1994  
 A;Title: Structural and functional aspects of the multiplicity of Neu differentiation factors.  
 A;Reference number: A56210; MUID:94158863; PMID:7509448  
 A;Accession: I38406  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: mRNA  
 A;Residues: 'A',95-418,'F',420-645 <RES>  
 A;Cross-references: EMBL:U02328; NID:g408406; PIDN:AAA19953.1; PID:g408407  
 C;Genetics:  
 A;Gene: GDB:HGL  
 A;Cross-references: GDB:132656; OMIM:142445  
 A;Map position: 8p22-8p11  
 C;Superfamily: human heregulin; EGF homology; immunoglobulin homology  
 C;Keywords: alternative splicing  
 F;182-221/Domain: EGF homology <EGF>

Query Match 16.3%; Score 285.5; DB 2; Length 645;  
 Best Local Similarity 32.9%; Pred. No. 5.8e-14;  
 Matches 70; Conservative 33; Mismatches 59; Indels 51; Gaps 7;

```

QY      126 GKNLKKEVGKILCTDCAT-----RPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRW 178
          || ||| | : | : ||: ||| | | | : || :: :|
Db      11 GKGKKKERGSGKKPESAAGSQSPALPPRLKEMKSQESAAGSKLVLRCESTSEYSSLRFKW 70

QY      179 FKDGKELNRS---RDIRIKYGNRKNRLQFNKVKVEDAGEYVCEAENILGKDTVGRGL- 234
          ||:| |||| ::||:| | : | | : | : |||:| : || | :
Db      71 FKNGNELNRKNKPQNIKIQQKPGK--SELRINKASLADSGEYMCKVISKLGNDASANIT 128

QY      235 -----YV-----NSVSTTLSSWSG--HARKCNETAKS 259
          || : | : | : | | | | | :
Db      129 IVESNEIITGMPASTEGAYVSSSPIRISVSTEGANTSSSTSTSTTGTSHLVKCAEKEKT 188

QY      260 YCVNGGVCYYIEGINQLS---CKCPNGFFGQRC 289
          : |||| | : :: : : | |||| | | ||
Db      189 FCVNGGECFMVKDLNPSRYLCKCPNEFTGDRC 221

```

RESULT 11  
 A45769

acetylcholine receptor synthesis stimulator ARIA-1 precursor - chicken

C;Species: Gallus gallus (chicken)  
 C;Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 08-Sep-2002  
 C;Accession: A45769  
 R;Falls, D.L.; Rosen, K.M.; Corfas, G.; Lane, W.S.; Fischbach, G.D.  
 Cell 72, 801-815, 1993  
 A;Title: ARIA, a protein that stimulates acetylcholine receptor synthesis, is a member of the neu ligand family.  
 A;Reference number: A45769; MUID:93201602; PMID:8453670  
 A;Accession: A45769  
 A;Status: preliminary  
 A;Molecule type: mRNA; protein  
 A;Residues: 1-602 <FAL>  
 A;Cross-references: GB:L11264; NID:g212603; PIDN:AAA49037.1; PID:g212604  
 A;Experimental source: brain  
 A;Note: sequence extracted from NCBI backbone (NCBIN:127787, NCBIP:127788)  
 C;Superfamily: human heregulin; EGF homology; immunoglobulin homology

Query Match 16.3%; Score 284.5; DB 2; Length 602;  
 Best Local Similarity 32.8%; Pred. No. 6.4e-14;  
 Matches 62; Conservative 34; Mismatches 72; Indels 21; Gaps 5;

```

Qy      109 TEQPLVFKTAFAPLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAA 168
          :| || : | | | : | | | | | | | | | | | | | | | | | | | | | |
Db      5  SEGPLQYSLAPTQTDVNS-----SYNTVPPKLKEMKNQEVAVGQKLVLRCETT 52

Qy      169 AGNPQPSYRWFKDGKEL---NRSRDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENIL 225
          : | : | : | | | : | | : | | : | | : | | | | | | : |
Db      53 SEYPALRFKWLKNGKEITKKNRPENVKIP-KKQKKYSELHIYRATLADAGEYACRVSSKL 111

Qy      226 GKDTVRGRLYVNSVSTTLSSWSG--HARKCNETAKSYCVNGGVCYYIEGI---NQLSCKC 280
          | | : : : : : | : | : | | | | | : | | | | | | | : |
Db      112 GNDSTKASVIITDTNATSTSTTGTSHLTKCDIKQKAFQCVNGGECYMKDLPNPPRYLCRC 171

Qy      281 PNGFFGQRC 289
          || | | | |
Db      172 PNEFTGDRC 180
  
```

# RESULT 12

I61722

neu differentiation factor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 08-Sep-2002

C;Accession: I61722

R;Wen, D.; Suggs, S.V.; Karunakaran, D.; Liu, N.; Cupples, R.L.; Luo, Y.; Janssen, A.M.; Ben-Baruch, N.; Trollinger, D.B.; Jacobsen, V.L.; Meng, S. Mol. Cell. Biol. 14, 1909-1919, 1994

A;Title: Structural and functional aspects of the multiplicity of Neu differentiation factors.

A;Reference number: A56210; MUID:94158863; PMID:7509448

A;Accession: I61722

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-662 <RES>

A;Cross-references: EMBL:U02322; NID:g408394; PIDN:AAA19947.1; PID:g408395

C;Superfamily: human heregulin; EGF homology; immunoglobulin homology

F;182-221/Domain: EGF homology <EGF>



QY 280 CPNGFFGQRC 289  
||| | | ||  
Db 201 CPNEFTGDRC 210

RESULT 14

I61718  
neu differentiation factor - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 08-Sep-2002  
C;Accession: I61718; I61721; I61720  
R;Wen, D.; Suggs, S.V.; Karunakaran, D.; Liu, N.; Cupples, R.L.; Luo, Y.;  
Janssen, A.M.; Ben-Baruch, N.; Trollinger, D.B.; Jacobsen, V.L.; Meng, S.  
Mol. Cell. Biol. 14, 1909-1919, 1994  
A;Title: Structural and functional aspects of the multiplicity of Neu  
differentiation factors.  
A;Reference number: A56210; MUID:94158863; PMID:7509448  
A;Accession: I61718  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-636 <RES>  
A;Cross-references: EMBL:U02318; NID:g408386; PIDN:AAA19943.1; PID:g408387  
A;Accession: I61721  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-444,'A',446-636 <RE2>  
A;Cross-references: EMBL:U02321; NID:g408392; PIDN:AAA19946.1; PID:g408393  
A;Accession: I61720  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-298,386,'V',388,'TR',391 <RE3>  
A;Cross-references: EMBL:U02320; NID:g408390; PIDN:AAA19945.1; PID:g408391  
C;Superfamily: human heregulin; EGF homology; immunoglobulin homology  
F;182-221/Domain: EGF homology <EGF>

Query Match 15.9%; Score 278; DB 2; Length 636;  
Best Local Similarity 33.7%; Pred. No. 2.1e-13;  
Matches 64; Conservative 31; Mismatches 51; Indels 44; Gaps 6;

QY 142 ATRPKLKKMKSQTGQVGKQSLKCEAAAGNPQPSYRWFKDGKELNRS---RDIRIKYGNG 198  
| :||:|||| | | :|| :: :|||:| ||| :||: |  
Db 34 ALPRLKEMKSQESAAGSKVLRCETSSEYSSLRFKWFKNELNRKNKPKENIKIQKKPG 93  
  
QY 199 RKNSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGL-----YV--- 236  
: | :| :| :| :|||:| : || | : : ||  
Db 94 K--SELRINKASLADSGEYMCKVISKLGNDASANITIVESNEFITGMPASTETAYVSSE 151  
  
QY 237 -----NSVSTTLSSWSG--HARKCNETAKSYCVNGGVCYYIEGINQLS---CK 279  
| :| :| :| :| | || | :||| || : : :| ||  
Db 152 SPIRISVSTEGANTSSSTSTSTTGTSHLIKCAEKEKTFVNGGECFTVKDLSNPSRYLCK 211  
  
QY 280 CPNGFFGQRC 289  
||| | | ||  
Db 212 CPNEFTGDRC 221

RESULT 15

S32359  
glial growth factor - bovine  
C;Species: Bos primigenius taurus (cattle)  
C;Date: 19-Mar-1997 #sequence\_revision 01-Aug-1997 #text\_change 08-Sep-2002  
C;Accession: S32359  
R;Marchionni, M.A.; Goodearl, A.D.J.; Chen, M.S.; Bermingham-McDonogh, O.; Kirk, C.; Hendricks, M.; Danehy, F.; Misumi, D.; Sudhalter, J.; Kobayashi, K.; Wroblewski, D.; Lynch, C.; Baldassare, M.; Hiles, I.; Davis, J.B.; Hsuan, J.J.; Totty, N.F.; Otsu, M.; McBurney, R.N.; Waterfield, M.D.; Stroobant, P.; Gwynne, D.  
Nature 362, 312-318, 1993  
A;Title: Glial growth factors are alternatively spliced erbB2 ligands expressed in the nervous system.  
A;Reference number: S32357; MUID:93205115; PMID:8096067  
A;Accession: S32359  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-241 <MAR>  
A;Cross-references: GB:L12259; NID:g289413; PIDN:AAA30540.1; PID:g289414  
C;Superfamily: human heregulin; EGF homology; immunoglobulin homology  
F;182-221/Domain: EGF homology <EGF>

Query Match 15.5%; Score 271; DB 2; Length 241;  
Best Local Similarity 32.6%; Pred. No. 2.3e-13;  
Matches 62; Conservative 34; Mismatches 50; Indels 44; Gaps 6;

```

Qy      142 ATRPKLKKMKSQTGQVGEEKQSLKCEAAAGNPQPSYRWFKDGKEL---NRSRDIRIKYGNG 198
      |  |::|:| | | |  |  |::|:|  :  :::|:| | |  |: ::|:|:  |
Db      34 ALPRLKEMKSQESVAGSKLVLRCEISSEYSSLKFKWFKNGSEL SRKNKPQNIKIQRPG 93

Qy      199 RKNSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGL-----YV--- 236
      :  | |: :|  : |::|:|:  : || |:  :
Db      94 K--SELRISKASLADSGEYMCKVISKLGNDASANITIVESNEITTGMPASTETAYVSSE 151

Qy      237 -----NSVSTTLSSWSG--HARKCNETAKSYCVNGGVCYYIEGINQLS---CK 279
      |: |:| :| :|  |  || |  |::| | | | |: :: :|  ||
Db      152 SPIRISVSTEGTNTSSSTSTSTAGTSHLVKCAEKEKTFVNGGECFMVKDLSNPSRYLCK 211

Qy      280 CPNGFFGQRC 289
      ||| | | |
Db      212 CPNEFTGDRC 221

```

Search completed: August 17, 2004, 14:13:19  
Job time : 16.2389 secs

OM protein - protein search, using sw model

Run on: August 17, 2004, 14:12:46 ; Search time 44.6656 Seconds  
 (without alignments)  
 2319.368 Million cell updates/sec

Title: US-09-864-675-2  
 Perfect score: 1749  
 Sequence: 1 MRRDPAPGFSMLLFGVSLAC.....PGTGVSSSQWSTSPSTLDLN 330

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 1292805 seqs, 313927144 residues

Total number of hits satisfying chosen parameters: 1292805

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result	Query	
No.	Score	Match Length DB ID Description

1	1749	100.0	330	9	US-09-864-675-2	Sequence 2, Appli
2	1749	100.0	422	15	US-10-447-839A-3	Sequence 3, Appli
3	1720	98.3	426	15	US-10-447-839A-2	Sequence 2, Appli
4	1610	92.1	850	16	US-10-408-765A-610	Sequence 610, App
5	1505	86.0	298	9	US-09-864-675-4	Sequence 4, Appli
6	960	54.9	181	13	US-10-096-241-4	Sequence 4, Appli
7	881	50.4	469	13	US-10-096-241-8	Sequence 8, Appli
8	881	50.4	647	13	US-10-096-241-32	Sequence 32, Appl
9	842	48.1	407	13	US-10-096-241-6	Sequence 6, Appli
10	821	46.9	605	13	US-10-096-241-2	Sequence 2, Appli
11	748	42.8	139	13	US-10-096-241-33	Sequence 33, Appl
12	523	29.9	422	8	US-08-736-019-170	Sequence 170, App
13	521	29.8	418	9	US-09-795-668-3	Sequence 3, Appli
14	521	29.8	418	9	US-09-795-686-3	Sequence 3, Appli
15	521	29.8	418	9	US-09-946-807-3	Sequence 3, Appli
16	354	20.2	204	9	US-09-795-668-4	Sequence 4, Appli
17	354	20.2	204	9	US-09-795-686-4	Sequence 4, Appli
18	354	20.2	204	9	US-09-946-807-4	Sequence 4, Appli
19	305	17.4	163	9	US-09-795-668-5	Sequence 5, Appli
20	305	17.4	163	9	US-09-795-686-5	Sequence 5, Appli
21	305	17.4	163	9	US-09-946-807-5	Sequence 5, Appli
22	304	17.4	501	14	US-10-290-578-10	Sequence 10, Appl
23	304	17.4	768	9	US-09-773-517-11	Sequence 11, Appl
24	304	17.4	768	9	US-09-792-025-11	Sequence 11, Appl
25	304	17.4	768	9	US-09-849-868-11	Sequence 11, Appl
26	304	17.4	768	9	US-09-808-602-85	Sequence 85, Appl
27	304	17.4	768	14	US-10-290-578-2	Sequence 2, Appli
28	304	17.4	768	14	US-10-453-183-11	Sequence 11, Appl
29	296	16.9	192	9	US-09-795-668-2	Sequence 2, Appli
30	296	16.9	192	9	US-09-795-686-2	Sequence 2, Appli
31	296	16.9	192	9	US-09-946-807-2	Sequence 2, Appli
32	293.5	16.8	456	9	US-09-795-668-17	Sequence 17, Appl
33	293.5	16.8	456	9	US-09-795-686-17	Sequence 17, Appl
34	293.5	16.8	456	9	US-09-946-807-17	Sequence 17, Appl
35	293.5	16.8	462	16	US-10-408-765A-883	Sequence 883, App
36	293.5	16.8	632	9	US-09-795-668-16	Sequence 16, Appl
37	293.5	16.8	632	9	US-09-795-686-16	Sequence 16, Appl
38	293.5	16.8	632	9	US-09-946-807-16	Sequence 16, Appl
39	293.5	16.8	669	9	US-09-773-517-1	Sequence 1, Appli
40	293.5	16.8	669	9	US-09-792-025-1	Sequence 1, Appli
41	293.5	16.8	669	9	US-09-849-868-1	Sequence 1, Appli
42	293.5	16.8	669	14	US-10-022-609-11	Sequence 11, Appl
43	293.5	16.8	669	14	US-10-453-183-1	Sequence 1, Appli
44	286	16.4	422	13	US-10-096-241-9	Sequence 9, Appli
45	285.5	16.3	239	9	US-09-795-668-18	Sequence 18, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-864-675-2

; Sequence 2, Application US/09864675

; Patent No. US20020081286A1

; GENERAL INFORMATION:

; APPLICANT: Marchionni, Mark

```

; TITLE OF INVENTION: NRG-2 NUCLEIC ACID MOLECULES,
; TITLE OF INVENTION: POLYPEPTIDES, AND DIAGNOSTIC AND THERAPEUTIC METHODS
; FILE REFERENCE: 04585/049002
; CURRENT APPLICATION NUMBER: US/09/864,675
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/206,495
; PRIOR FILING DATE: 2000-05-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-675-2

```

```

Query Match          100.0%; Score 1749; DB 9; Length 330;
Best Local Similarity 100.0%; Pred. No. 6.5e-135;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60

Qy     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 120

Qy    121 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGGVGEKQSLKCEAAAGNPQPSYRWFK 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGGVGEKQSLKCEAAAGNPQPSYRWFK 180

Qy    181 DGKELNRSRDIRIKYGNRKN SRLQFNKVVEDAGEYVCEAENILGKDTVGRGRLYVNSVS 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 DGKELNRSRDIRIKYGNRKN SRLQFNKVVEDAGEYVCEAENILGKDTVGRGRLYVNSVS 240

Qy    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP 300

Qy    301 DPKQSVLWDTPGTGVSSSQWSTSPSTLDLN 330
        ||||||||||||||||||||||||||||||||
Db    301 DPKQSVLWDTPGTGVSSSQWSTSPSTLDLN 330

```

# RESULT 2

US-10-447-839A-3

```

; Sequence 3, Application US/10447839A
; Publication No. US20040018181A1

```

## GENERAL INFORMATION:

```

; APPLICANT: Kufe, Donald W.
; APPLICANT: Kharbanda, Surender
; APPLICANT: Weitman, Steven D.

```

; TITLE OF INVENTION: MUC1 INTERFERENCE RNA COMPOSITIONS AND METHODS DERIVED THEREFROM

```

; FILE REFERENCE: 1000.05.009
; CURRENT APPLICATION NUMBER: US/10/447,839A
; CURRENT FILING DATE: 2003-05-29

```



; PRIOR APPLICATION NUMBER: 10/293,391  
 ; PRIOR FILING DATE: 2002-11-13  
 ; PRIOR APPLICATION NUMBER: 09/951,938  
 ; PRIOR FILING DATE: 2001-09-11  
 ; PRIOR APPLICATION NUMBER: 60/231,841  
 ; PRIOR FILING DATE: 2000-09-11  
 ; NUMBER OF SEQ ID NOS: 109  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 3  
 ; LENGTH: 422  
 ; TYPE: PRT  
 ; ORGANISM: Homo Sapien  
 US-10-447-839A-3

Query Match 100.0%; Score 1749; DB 15; Length 422;  
 Best Local Similarity 100.0%; Pred. No. 8.9e-135;  
 Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP	60
Db	93	MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP	152
QY	61	PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA	120
Db	153	PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA	212
QY	121	PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK	180
Db	213	PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK	272
QY	181	DGKELNRSRDIRIKYGNRKN SRLQFNKVVEDAGEYVCEAENILGKDTVGRRLYVNSVS	240
Db	273	DGKELNRSRDIRIKYGNRKN SRLQFNKVVEDAGEYVCEAENILGKDTVGRRLYVNSVS	332
QY	241	TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP	300
Db	333	TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP	392
QY	301	DPKQSVLWDTPGTGVSSSQWSTSPSTLDLN	330
Db	393	DPKQSVLWDTPGTGVSSSQWSTSPSTLDLN	422

# RESULT 3

US-10-447-839A-2

; Sequence 2, Application US/10447839A

; Publication No. US20040018181A1

## ; GENERAL INFORMATION:

; APPLICANT: Kufe, Donald W.

; APPLICANT: Kharbanda, Surender

; APPLICANT: Weitman, Steven D.

; TITLE OF INVENTION: MUC1 INTERFERENCE RNA COMPOSITIONS AND METHODS DERIVED THEREFROM

; FILE REFERENCE: 1000.05.009

; CURRENT APPLICATION NUMBER: US/10/447,839A

; CURRENT FILING DATE: 2003-05-29

; PRIOR APPLICATION NUMBER: 10/293,391

; PRIOR FILING DATE: 2002-11-13  
 ; PRIOR APPLICATION NUMBER: 09/951,938  
 ; PRIOR FILING DATE: 2001-09-11  
 ; PRIOR APPLICATION NUMBER: 60/231,841  
 ; PRIOR FILING DATE: 2000-09-11  
 ; NUMBER OF SEQ ID NOS: 109  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 2  
 ; LENGTH: 426  
 ; TYPE: PRT  
 ; ORGANISM: Homo Sapien  
 US-10-447-839A-2

Query Match 98.3%; Score 1720; DB 15; Length 426;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-132;  
 Matches 324; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MRRDPAPGFSMLLFGVSLACYSPSLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP	60
Db	93	MRRDPAPGFSMLLFGVSLACYSPSLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP	152
Qy	61	PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAF	120
Db	153	PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAF	212
Qy	121	PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGGQVGEKQSLKCEAAAGNPQPSYRWFK	180
Db	213	PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGGQVGEKQSLKCEAAAGNPQPSYRWFK	272
Qy	181	DGKELNRSRDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS	240
Db	273	DGKELNRSRDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS	332
Qy	241	TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP	300
Db	333	TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP	392
Qy	301	DPKQSVLWDTPGTGVSSSQWSTSP	324
Db	393	DPKQSVLWDTPGTGVSSSQWSTSP	416

#### RESULT 4

US-10-408-765A-610

; Sequence 610, Application US/10408765A  
 ; Publication No. US20040101874A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ghosh, Soumitra S.  
 ; APPLICANT: Fahy, Eoin D.  
 ; APPLICANT: Zhang, Bing  
 ; APPLICANT: Gibson, Bradford W.  
 ; APPLICANT: Taylor, Steven W.  
 ; APPLICANT: Glenn, Gary M.  
 ; APPLICANT: Warnock, Dale E.  
 ; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
 ; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME  
 ; FILE REFERENCE: 660088.465

; CURRENT APPLICATION NUMBER: US/10/408,765A  
; CURRENT FILING DATE: 2003-04-04  
; NUMBER OF SEQ ID NOS: 3077  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 610  
; LENGTH: 850  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-408-765A-610

Query Match 92.1%; Score 1610; DB 16; Length 850;  
Best Local Similarity 100.0%; Pred. No. 5.3e-123;  
Matches 304; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
      |||
Db      93 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 152

QY     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 120
      |||
Db    153 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 212

QY    121 PLDTNGKNLKKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180
      |||
Db    213 PLDTNGKNLKKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 272

QY    181 DGKELNRSRDIRIKYGNRKNSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240
      |||
Db    273 DGKELNRSRDIRIKYGNRKNSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS 332

QY    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP 300
      |||
Db    333 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP 392

QY    301 DPKQ 304
      |||
Db    393 DPKQ 396
```

RESULT 5

US-09-864-675-4

; Sequence 4, Application US/09864675  
; Patent No. US20020081286A1  
; GENERAL INFORMATION:  
; APPLICANT: Marchionni, Mark  
; TITLE OF INVENTION: NRG-2 NUCLEIC ACID MOLECULES,  
; TITLE OF INVENTION: POLYPEPTIDES, AND DIAGNOSTIC AND THERAPEUTIC METHODS  
; FILE REFERENCE: 04585/049002  
; CURRENT APPLICATION NUMBER: US/09/864,675  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/206,495  
; PRIOR FILING DATE: 2000-05-23  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 298  
; TYPE: PRT

; ORGANISM: Homo sapiens  
US-09-864-675-4

Query Match 86.0%; Score 1505; DB 9; Length 298;  
Best Local Similarity 98.6%; Pred. No. 5.5e-115;  
Matches 285; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
QY      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
          |||
Db      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60

QY     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFE 120
          |||
Db     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFE 120

QY    121 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180
          |||
Db    121 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180

QY    181 DGKELNRSRDIRIKYGNGRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240
          |||
Db    181 DGKELNRSRDIRIKYGNGRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240

QY    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRC 289
          |||
Db    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDR 289
```

RESULT 6

US-10-096-241-4

; Sequence 4, Application US/10096241  
; Publication No. US20020127594A1

; GENERAL INFORMATION:

```
; APPLICANT: Gearing, David P.
;           Busfield, Samantha J.
; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
;                   AND USES THEREFOR
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson P.C.
;   STREET: 225 Franklin Street
;   CITY: Boston
;   STATE: MA
;   COUNTRY: US
;   ZIP: 02110-2804
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/096,241
;   FILING DATE: 12-Mar-2002
;   CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/699,591
;   FILING DATE: 19-AUG-1996
```

```

;      ATTORNEY/AGENT INFORMATION:
;      NAME: Fasse, J. Peter
;      REGISTRATION NUMBER: 32,983
;      REFERENCE/DOCKET NUMBER: 07334/022001
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE: 617-542-5070
;      TELEFAX: 617-542-8906
;      TELEX: <Unknown>
;      INFORMATION FOR SEQ ID NO: 4:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 181 amino acids
;      TYPE: amino acid
;      STRANDEDNESS: not relevant
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
;      FRAGMENT TYPE: internal
;      SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-096-241-4

```

```

Query Match          54.9%;  Score 960;  DB 13;  Length 181;
Best Local Similarity 97.8%;  Pred. No. 1.2e-70;
Matches 177;  Conservative 3;  Mismatches 1;  Indels 0;  Gaps 0;

```

```

Qy      150 MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKNLSRLQFNKV 209
        |||||:|||||
Db      1   MKSQTGEVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNVRKNLSRLQFNKV 60

Qy      210 KVEDAGEYVCEAENILGKDTVRGRLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 269
        :|||||:|||||
Db      61 RVEDAGEYVCEAENILGKDTVRGRHLVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 120

Qy      270 IEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQSVLWDTPGTGVSSSQWSTSPSTLDL 329
        |||||
Db      121 IEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQSVLWDTPGTGVSSSQWSTSPSTLDL 180

Qy      330 N 330
        |
Db      181 N 181

```

# RESULT 7

US-10-096-241-8

; Sequence 8, Application US/10096241

; Publication No. US20020127594A1

## ; GENERAL INFORMATION:

```

;      APPLICANT: Gearing, David P.
;      Busfield, Samantha J.
;      TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
;      AND USES THEREFOR
;      NUMBER OF SEQUENCES: 33
;      CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Fish & Richardson P.C.
;      STREET: 225 Franklin Street
;      CITY: Boston
;      STATE: MA
;      COUNTRY: US
;      ZIP: 02110-2804

```

```

;      COMPUTER READABLE FORM:
;      MEDIUM TYPE: Diskette
;      COMPUTER: IBM Compatible
;      OPERATING SYSTEM: DOS
;      SOFTWARE: FastSEQ Version 2.0
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/10/096,241
;      FILING DATE: 12-Mar-2002
;      CLASSIFICATION: <Unknown>
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 08/699,591
;      FILING DATE: 19-AUG-1996
;      ATTORNEY/AGENT INFORMATION:
;      NAME: Fasse, J. Peter
;      REGISTRATION NUMBER: 32,983
;      REFERENCE/DOCKET NUMBER: 07334/022001
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE: 617-542-5070
;      TELEFAX: 617-542-8906
;      TELEX: <Unknown>
;      INFORMATION FOR SEQ ID NO: 8:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 469 amino acids
;      TYPE: amino acid
;      STRANDEDNESS: not relevant
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
;      FRAGMENT TYPE: internal
;      SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-096-241-8

```

```

Query Match          50.4%;  Score 881;  DB 13;  Length 469;
Best Local Similarity 100.0%;  Pred. No. 1.2e-63;
Matches 163;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

Qy      142 ATRPKLKKMKSQTGQVGEEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNGRKN 201
          |||
Db      31 ATRPKLKKMKSQTGQVGEEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNGRKN 90

Qy      202 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLVNSVSTTLSSWSGHARKCNETAKSYC 261
          |||
Db      91 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLVNSVSTTLSSWSGHARKCNETAKSYC 150

Qy      262 VNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ 304
          |||
Db      151 VNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ 193

```

# RESULT 8

US-10-096-241-32

; Sequence 32, Application US/10096241

; Publication No. US20020127594A1

## ; GENERAL INFORMATION:

; APPLICANT: Gearing, David P.

; Busfield, Samantha J.

; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES  
; AND USES THEREFOR

```

;      NUMBER OF SEQUENCES: 33
;      CORRESPONDENCE ADDRESS:
;          ADDRESSEE: Fish & Richardson P.C.
;          STREET: 225 Franklin Street
;          CITY: Boston
;          STATE: MA
;          COUNTRY: US
;          ZIP: 02110-2804
;      COMPUTER READABLE FORM:
;          MEDIUM TYPE: Diskette
;          COMPUTER: IBM Compatible
;          OPERATING SYSTEM: DOS
;          SOFTWARE: FastSEQ Version 2.0
;      CURRENT APPLICATION DATA:
;          APPLICATION NUMBER: US/10/096,241
;          FILING DATE: 12-Mar-2002
;          CLASSIFICATION: <Unknown>
;      PRIOR APPLICATION DATA:
;          APPLICATION NUMBER: 08/699,591
;          FILING DATE: 19-AUG-1996
;      ATTORNEY/AGENT INFORMATION:
;          NAME: Fasse, J. Peter
;          REGISTRATION NUMBER: 32,983
;          REFERENCE/DOCKET NUMBER: 07334/022001
;      TELECOMMUNICATION INFORMATION:
;          TELEPHONE: 617-542-5070
;          TELEFAX: 617-542-8906
;          TELEX: <Unknown>
;      INFORMATION FOR SEQ ID NO: 32:
;          SEQUENCE CHARACTERISTICS:
;              LENGTH: 647 amino acids
;              TYPE: amino acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;          MOLECULE TYPE: protein
;          FRAGMENT TYPE: internal
;          SEQUENCE DESCRIPTION: SEQ ID NO: 32:
US-10-096-241-32

```

```

Query Match          50.4%;  Score 881;  DB 13;  Length 647;
Best Local Similarity 100.0%;  Pred. No. 1.9e-63;
Matches 163;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

Qy      142 ATRPKLKKMKSQTGQVG EKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNGRKN 201
          |||
Db      31 ATRPKLKKMKSQTGQVG EKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNGRKN 90

Qy      202 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC 261
          |||
Db      91 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC 150

Qy      262 VNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ 304
          |||
Db      151 VNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQ 193

```

RESULT 9

US-10-096-241-6

```

; Sequence 6, Application US/10096241
; Publication No. US20020127594A1
; GENERAL INFORMATION:
;   APPLICANT: Gearing, David P.
;               Busfield, Samantha J.
;   TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
;                       AND USES THEREFOR
;   NUMBER OF SEQUENCES: 33
;   CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Fish & Richardson P.C.
;       STREET: 225 Franklin Street
;       CITY: Boston
;       STATE: MA
;       COUNTRY: US
;       ZIP: 02110-2804
;   COMPUTER READABLE FORM:
;       MEDIUM TYPE: Diskette
;       COMPUTER: IBM Compatible
;       OPERATING SYSTEM: DOS
;       SOFTWARE: FastSEQ Version 2.0
;   CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/10/096,241
;       FILING DATE: 12-Mar-2002
;       CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 08/699,591
;       FILING DATE: 19-AUG-1996
;   ATTORNEY/AGENT INFORMATION:
;       NAME: Fasse, J. Peter
;       REGISTRATION NUMBER: 32,983
;       REFERENCE/DOCKET NUMBER: 07334/022001
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 617-542-5070
;       TELEFAX: 617-542-8906
;       TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 6:
;       SEQUENCE CHARACTERISTICS:
;           LENGTH: 407 amino acids
;           TYPE: amino acid
;           STRANDEDNESS: not relevant
;           TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       FRAGMENT TYPE: internal
;       SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```

US-10-096-241-6

```

Query Match          48.1%; Score 842; DB 13; Length 407;
Best Local Similarity 98.7%; Pred. No. 1.6e-60;
Matches 156; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      150 MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNGRKNSRLQFNKV 209
          |||
Db       1 MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNGRKNSRLQFNKV 60
          |||
QY      210 KVEDAGEYVCEAENILGKDTVGRGLVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 269
          |||

```



Db 61 KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 120

Qy 270 IEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQSVL 307  
 |||

Db 121 IEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQKHL 158

RESULT 10  
 US-10-096-241-2  
 ; Sequence 2, Application US/10096241  
 ; Publication No. US20020127594A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gearing, David P.  
 ; Busfield, Samantha J.  
 ; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES  
 ; AND USES THEREFOR  
 ; NUMBER OF SEQUENCES: 33  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Richardson P.C.  
 ; STREET: 225 Franklin Street  
 ; CITY: Boston  
 ; STATE: MA  
 ; COUNTRY: US  
 ; ZIP: 02110-2804  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSEQ Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/096,241  
 ; FILING DATE: 12-Mar-2002  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/699,591  
 ; FILING DATE: 19-AUG-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Fasse, J. Peter  
 ; REGISTRATION NUMBER: 32,983  
 ; REFERENCE/DOCKET NUMBER: 07334/022001  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-542-5070  
 ; TELEFAX: 617-542-8906  
 ; TELEX: <Unknown>  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 605 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: not relevant  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; FRAGMENT TYPE: internal  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
 US-10-096-241-2

Query Match 46.9%; Score 821; DB 13; Length 605;  
 Best Local Similarity 97.4%; Pred. No. 1.4e-58;

Matches 151; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

RESULT 11

; Sequence 33, Application US/10096241

; GENERAL INFORMATION:

```

;          NUMBER OF SEQUENCES: 33

```

ADDRESSEE: Fish & Richardson P.C.

; CITY: Boston

; COUNTRY: US

; COMPUTER READABLE FO

```

;
COMPUTER: IBM Compatil

```

```

; SOFTWARE: FastSEO Ver

```

APPLICATION NUMBER:

CLASSIFICATION: <Unknown

APPLICATION NUMBER

ATTORNEY/AGENT INFORMATION:

REGISTRATION NUMBER:

TELECOMMUNICATION INFORMATION:

TELEFAX: 617-542-8906

INFORMATION FOR SEO ID NO:

LENGTH: 139 amino acids

STRANDEDNESS: not relevant

; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 33:  
US-10-096-241-33

Query Match 42.8%; Score 748; DB 13; Length 139;  
Best Local Similarity 98.6%; Pred. No. 2e-53;  
Matches 137; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 192 RIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHAR 251  
| | | | | : | | | | | : | | | | |  
Db 1 RIKYGNRKNRSLQFNKVRVEDAGEYVCEAENILGKDTVGRGLHVNSVSTTLSSWSGHAR 60  
  
QY 252 KCNETAKSYCVNGGVVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQSVLWDTP 311  
| | | | | : | | | | | : | | | | |  
Db 61 KCNETAKSYCVNGGVVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQSVLWDTP 120  
  
QY 312 GTGVSSSQWSTSPSTLDLN 330  
| | | | |  
Db 121 GTGVSSSQWSTSPSTLDLN 139

RESULT 12

US-08-736-019-170

; Sequence 170, Application US/08736019

; Publication No. US20030207799A1

; GENERAL INFORMATION:

; APPLICANT: Goodearl, Andrew

; APPLICANT: Stroobant, Paul

; APPLICANT: Minghetti, Luisa

; APPLICANT: Waterfield, Michael

; APPLICANT: Marchionni, Mark

; APPLICANT: Chen, Mario

; APPLICANT: Hiles, Ian

; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR

; TITLE OF INVENTION: PREPARATION AND USE

; NUMBER OF SEQUENCES: 189

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Clark & Elbing LLP

; STREET: 176 Federal Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02110

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; COMPUTER: IBM Compatible Pentium

; OPERATING SYSTEM: Windows95

; SOFTWARE: FastSeq Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/736,019

; FILING DATE: 22-OCT-1996

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/471,833

; FILING DATE: 06-JUN-1995

; PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: 08/036,555
; FILING DATE: 24-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/965,173
; FILING DATE: 23-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/907,138
; FILING DATE: 30-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/940,389
; FILING DATE: 03-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/863,703
; FILING DATE: 03-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 91 07566.3
; FILING DATE: 10-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bieker-Brady, Kristina
; REGISTRATION NUMBER: 39,109
; REFERENCE/DOCKET NUMBER: 04585/00200Q
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 428-0200
; TELEFAX: (617) 428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 170:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-736-019-170

```

```

Query Match          29.9%; Score 523; DB 8; Length 422;
Best Local Similarity 35.5%; Pred. No. 2.3e-34;
Matches 124; Conservative 59; Mismatches 88; Indels 78; Gaps 13;

```

```

Qy      15 GVSLACYS--PSLKSVQDQAYKAPVVVEGKV-----QGLV-----PAGGSSS--NSTRE 59
      | |: ||| ||: |||: | :| ||: ||| | | : | | : | |
Db      58 GASV-CYSSPPSVGSVQELAQRAAVVIEGKVHPQRRQQGALDRKAAAAAGEAGAWGGDRE 116

Qy      60 PPASGRVA-----LVKVLDDKWPLRSGGLQ 83
      |||:| | | | | | | | | | | | | | | | | | | | | |
Db     117 PPAAGPRALGPPAEPELLAANGTVPSWPTAPVPSAGEPGEEAPYLVKVHQVWAVKAGGLK 176

Qy      84 REQVISV-----GSCVPLERNQRYIFFLEP-----TEQPLVFKTAFAPLDTNGKN 128
      :: ::| | | | : : | | | | : | | : : | | | | : |
Db     177 KDSLLTVRLGTWGHFAFPSCGRLKEDSRYIFFMEPDANSTSRAPAAFRASFPPLET-GRN 235

Qy     129 LKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRS 188
      ||||| ::| | | | : ||| | | | | : | | : | | | | | |
Db     236 LKKEVSRVLCKRCALPPQLKEMKSQESAAGSKLVLRCESSSEYSSSLRFKWFKNGNELNRK 295

Qy     189 ---RDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSS 245
      ::|:| | : | | : | | : | | | : | | : | | : | | : |
Db     296 NKPQNIKIQKKPGK--SELRINKASLADSGEYMCKVISKLGNDSSANITIVESNATSTS 353

```

Qy 246 WSG--HARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPNGFFGQRC 289  
:| | || | |:|||| |: :: :| | |||| | ||  
Db 354 TTGTSHLVKCAEKEKTFVCVNGGECFMVKDLSNPSRYLCKCPNEFTGDRC 402

RESULT 13

US-09-795-668-3  
; Sequence 3, Application US/09795668  
; Patent No. US20020045577A1  
; GENERAL INFORMATION:  
; APPLICANT: Stefansson, Hreinn  
; APPLICANT: Steinthorsdottir, Valgerdur  
; APPLICANT: Gulcher, Jeffrey R.  
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE  
; FILE REFERENCE: 2345.2004-001  
; CURRENT APPLICATION NUMBER: US/09/795,668  
; CURRENT FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 09/515,716  
; PRIOR FILING DATE: 2000-02-28  
; NUMBER OF SEQ ID NOS: 1531  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 418  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-795-668-3

Query Match 29.8%; Score 521; DB 9; Length 418;  
Best Local Similarity 35.2%; Pred. No. 3.3e-34;  
Matches 122; Conservative 59; Mismatches 90; Indels 76; Gaps 12;

Qy 15 GVSLACYSPSLKSVQDQAYKAPVVVEGKV-----QGLV-----PAGGSSS--NSTREPP 61  
| |: | |: ||: |: |: ||: ||| || : | | : ||||  
Db 56 GASV-CSPPSVGSVQELAQRAAVVIEGKVHPQRRQQGALDRKAAAAAGEAGAWGGDREPP 114  
  
Qy 62 ASGRVA-----LVKVLDKWPLRSGGLQRE 85  
|:| | ||| | :|||:|:  
Db 115 AAGPRALGPPAAEEPLLAANGTVPSWPTAPVPSAGEPGEEAPYLVKVHVQVWAVKAGGLKKD 174  
  
Qy 86 QVISV-----GSCVPLERNQRYIFFLEP-----TEQPLVFKTAFAPLDTNGKNLK 130  
:::| || |: : ||||:| | : | |: :| |:| |:|||  
Db 175 SLLTVRLGTWGHPAFPSCGRLKEDSRYIFFMEPDANSTSRAPAAFRASFPPLET-GRNLK 233  
  
Qy 131 KEVGKILCTDCATRPKLKMKMSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGGKELNRS-- 188  
||| ::|| || |:||:|||| | | |:|| :: :|||:| ||||  
Db 234 KEVSRVLCKRCALPPRLKEMKSQESAAGSKLVLRCESSSEYSSLRFKWFKNGNELNRKNK 293  
  
Qy 189 -RDIRIKYGNGRKNSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWS 247  
:::|: | : | | : |:|||:| : || |: : : :| :| :  
Db 294 PQNIKIQQKPGK--SELRINKASLADSGEYMCKVISKLGNDASANITIVESNATSTSTT 351  
  
Qy 248 G--HARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPNGFFGQRC 289  
| | || | |:|||| |: :: :| | |||| | ||  
Db 352 GTSHLVKCAEKEKTFVCVNGGECFMVKDLSNPSRYLCKCPNEFTGDRC 398

RESULT 14

US-09-795-686-3  
; Sequence 3, Application US/09795686  
; Patent No. US20020094954A1  
; GENERAL INFORMATION:  
; APPLICANT: Stefansson, Hreinn  
; APPLICANT: Steinthorsdottir, Valgerdur  
; APPLICANT: Gulcher, Jeffrey R.  
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE  
; FILE REFERENCE: 2345.2005-001  
; CURRENT APPLICATION NUMBER: US/09/795,686  
; CURRENT FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 09/515,715  
; PRIOR FILING DATE: 2000-02-28  
; NUMBER OF SEQ ID NOS: 1531  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 418  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-795-686-3

Query Match 29.8%; Score 521; DB 9; Length 418;  
Best Local Similarity 35.2%; Pred. No. 3.3e-34;  
Matches 122; Conservative 59; Mismatches 90; Indels 76; Gaps 12;

Qy	15	GVSLACYSPLKSVQDQAYKAPVVVEGKV-----QGLV-----PAGGSSS--NSTREPP	61
		:      :    :   :    :         :     :	
Db	56	GASV-CSPPSVGSVQELAQRAAVVIEGKVHPQRRQQGALDRKAAAAAGEAGAWGGDREPP	114
Qy	62	ASGRVA-----LVKVLDKWPLRSGGLQRE	85
		:          :   ::	
Db	115	AAGPRALGPPAEELLAAANGTVPSWPTAPVPSAGEPEGEEAPYLVKVHQVWAVKAGGLKKD	174
Qy	86	QVISV-----GSCVPLERNQRYIFFLEP-----TEQPLVFKTAFAPLDTNGKNLK	130
		:::       : :     :    :     : :    :	
Db	175	SLLTVRLGTWGHPAFPSCGRLKEDSRYIFFMEPDANSTSRAPAAFRASFPPLET-GRNLK	233
Qy	131	KEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGGKELNRS--	188
		::        :  :           :   :: : :   :	
Db	234	KEVSRVLCKRCALPPRLKEMKSQESAAGSKLVLRCETSSEYSSLRFKWFKNGNELNRKNK	293
Qy	189	-RDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWS	247
		:: : :   :     :    :   :   : : :      : : :   :  :	
Db	294	PQNIKIQKKPGK--SELRINKASLADSGEYMCKVISKLGNDASANITIVESNATSTSTT	351
Qy	248	G--HARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPNGFFGQRC	289
		:         : : :	
Db	352	GTSHLVKCAEKEKTFCVNGGECFMVKDL SNPSRYLCKCPNEFTGDRC	398

# RESULT 15

US-09-946-807-3  
; Sequence 3, Application US/09946807  
; Patent No. US20020165144A1  
; GENERAL INFORMATION:  
; APPLICANT: Stefansson, Hreinn  
; APPLICANT: Steinthorsdottir, Valgerdur

```
; APPLICANT:  Gulcher, Jeffrey R.
; TITLE OF INVENTION:  HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE:  2345.2004-001
; CURRENT APPLICATION NUMBER:  US/09/946,807
; CURRENT FILING DATE:  2001-09-05
; PRIOR APPLICATION NUMBER:  US/09/795,668
; PRIOR FILING DATE:  2001-02-28
; PRIOR APPLICATION NUMBER:  US 09/515,716
; PRIOR FILING DATE:  2000-02-28
; NUMBER OF SEQ ID NOS:  1531
; SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH:  418
; TYPE:  PRT
; ORGANISM:  Homo sapiens
US-09-946-807-3
```

```
Query Match          29.8%;  Score 521;  DB 9;  Length 418;
Best Local Similarity  35.2%;  Pred. No. 3.3e-34;
Matches 122;  Conservative  59;  Mismatches  90;  Indels  76;  Gaps  12;
```

```
QY      15 GVSLACYSPSLKSVQDQAYKAPVVVEGKV-----QGLV-----PAGGSSS--NSTREPP 61
      | |: |  ||: |||: | :| ||: |||  || :  | | :  |||
Db      56 GASV-CSPPSVGSGVQELAQRAAVVIEGKVHPQRRQQGALDRKAAAAAGEAGAWGGDREPP 114

QY      62 ASGRVA-----LVKVLDKWPLRSGLQRE 85
      |:| |  |||  | ::|||:::
Db     115 AAGPRALGPPAEPELLAANGTVPSWPTAPVPSAGEPGEAPYLKVHVQVWAVKAGGLKKD 174

QY      86 QVISV-----GSCVPLERNQRYIFFLEP-----TEQPLVFKTAFAPLDTNGKNLK 130
      :::|  || | : : ||||: ||  : | | : : | ||: | :|||
Db     175 SLLTVRLGTWGHPAFPSCGRLKEDSRYIFFMEPDANSTSRAPAAFRASFPLET-GRNLK 233

QY     131 KEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRS-- 188
      ||| :|| | | | :||: ||||  | | | :|| : : |||: | |||
Db     234 KEVSRVLCKRCALPPRLKEMKSQESAAGSKLVLRCESTSSEYSSLRFKWKNGNELNRKNK 293

QY     189 -RDIRIKYGNGRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWS 247
      ::||: | : | | : || : | :|||: | : : || | :  : :  : | : | :
Db     294 PQNIKIQQKPGK--SELRINKASLADSGEYMCKVISKLGNDSASANITIVESNATSTSTT 351

QY     248 G--HARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPNGFFGQRC 289
      | | || | | : : |||| | : : : : |  |||| | | ||
Db     352 GTSHLVKCAEKEKTFCVNGGECFMVKDLSNPSRYLCKCPNEFTGDRC 398
```

```
Search completed: August 17, 2004, 14:22:29
Job time : 45.6656 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 17, 2004, 14:05:35 ; Search time 37.3089 Seconds  
(without alignments)  
2790.781 Million cell updates/sec

Title: US-09-864-675-2  
Perfect score: 1749  
Sequence: 1 MRRDPAPGFSMLLFGVSLAC.....PGTGVSSSQWSTSPSTLDLN 330

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

%  
Result Query  
No. Score Match Length DB ID Description  
-----



1	485	27.7	782	11	Q9ESA5	Q9esa5 rattus norv
2	460	26.3	342	11	Q9ESA1	Q9esa1 rattus norv
3	452	25.8	323	11	Q9ESA2	Q9esa2 rattus norv
4	450	25.7	317	11	Q9ESA3	Q9esa3 rattus norv
5	444	25.4	79	11	Q810X2	Q810x2 mus musculu
6	404.5	23.1	348	4	Q8NFN3	Q8nfn3 homo sapien
7	290	16.6	461	11	O35947	O35947 mesocricetu
8	271	15.5	241	6	Q07112	Q07112 bos taurus
9	234	13.4	211	11	Q8BKI8	Q8bki8 mus musculu
10	221	12.6	244	11	Q9ESA4	Q9esa4 rattus norv
11	200	11.4	54	11	Q810X1	Q810x1 mus musculu
12	180.5	10.3	167	4	Q8NFN2	Q8nfn2 homo sapien
13	166	9.5	8625	5	Q86GD6	Q86gd6 procambarus
14	164.5	9.4	5175	5	Q8I0L3	Q8i0l3 caenorhabdi
15	164.5	9.4	5198	5	O76518	O76518 caenorhabdi
16	158	9.0	1106	4	Q8WX93	Q8wx93 homo sapien
17	153	8.7	8943	5	Q9V4F7	Q9v4f7 drosophila
18	151	8.6	298	11	Q9JI59	Q9ji59 mus musculu
19	151	8.6	298	11	Q8C5K9	Q8c5k9 mus musculu
20	151	8.6	410	4	Q8N1M2	Q8n1m2 homo sapien
21	151	8.6	1323	13	Q08476	Q08476 gallus gall
22	150.5	8.6	6658	5	O76281	O76281 drosophila
23	150	8.6	298	11	Q8CE95	Q8ce95 mus musculu
24	150	8.6	3950	6	Q7YRF5	Q7yrf5 canis famil
25	149	8.5	512	4	Q96DN8	Q96dn8 homo sapien
26	149	8.5	5636	4	Q96RW7	Q96rw7 homo sapien
27	146.5	8.4	330	13	Q90Z42	Q90z42 gallus gall
28	146.5	8.4	4076	11	Q7TN00	Q7tn00 rattus norv
29	144.5	8.3	754	11	Q8BZ76	Q8bz76 mus musculu
30	144.5	8.3	858	5	O18466	O18466 hirudo medi
31	144.5	8.3	7962	4	Q10465	Q10465 homo sapien
32	144.5	8.3	34350	4	Q8WZ42	Q8wz42 homo sapien
33	144	8.2	338	4	Q8IV49	Q8iv49 homo sapien
34	143.5	8.2	507	4	Q96K90	Q96k90 homo sapien
35	143.5	8.2	1320	4	Q96KF5	Q96kf5 homo sapien
36	143.5	8.2	1320	4	Q86TC9	Q86tc9 homo sapien
37	143.5	8.2	1391	4	Q8N3L4	Q8n3l4 homo sapien
38	143.5	8.2	1612	11	O89026	O89026 mus musculu
39	143.5	8.2	1651	4	Q9Y6N7	Q9y6n7 homo sapien
40	142.5	8.1	947	5	Q26262	Q26262 caenorhabdi
41	142.5	8.1	947	5	O44171	O44171 caenorhabdi
42	142	8.1	2693	5	Q8ISF3	Q8isf3 caenorhabdi
43	142	8.1	2708	5	Q8ISF4	Q8isf4 caenorhabdi
44	142	8.1	2780	5	Q8MNS0	Q8mns0 caenorhabdi
45	142	8.1	2808	5	Q8MNS1	Q8mns1 caenorhabdi

# ALIGNMENTS

## RESULT 1

Q9ESA5

ID Q9ESA5 PRELIMINARY; PRT; 782 AA.

AC Q9ESA5;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Glial growth factor beta 1a (Fragment).  
 GN NRG1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord, and Brain stem;  
 RA Carroll S.L., Stonecypher M.S., Anderson K.D., Pearson R.J. Jr.,  
 RA Frohnert P.W.;  
 RT "Structural and Functional Diversity of Glial Growth Factor Isoforms  
 RT Expressed in Regenerating Peripheral Nerve and Associated Neurons.";  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF194993; AAG28433.1; -.  
 DR HSSP; Q12784; 1HRE.  
 DR GO; GO:0005102; F:receptor binding; IEA.  
 DR GO; GO:0005351; F:sugar porter activity; IEA.  
 DR GO; GO:0009790; P:embryonic development; IEA.  
 DR GO; GO:0009401; P:phosphoenolpyruvate-dependent sugar phospho. . .; IEA.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR002114; HPr\_SerP\_S.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR InterPro; IPR002154; Neuregulin.  
 DR Pfam; PF00008; EGF; 1.  
 DR Pfam; PF00047; ig; 1.  
 DR Pfam; PF02158; Neuregulin; 1.  
 DR PRINTS; PR01089; NEUREGULIN.  
 DR SMART; SM00181; EGF; 1.  
 DR SMART; SM00408; IGc2; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 DR PROSITE; PS00589; PTS\_HPR\_SER; 1.  
 KW EGF-like domain; Immunoglobulin domain.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 782 AA; 86036 MW; F6174A68F4E27BDE CRC64;

Query Match 27.7%; Score 485; DB 11; Length 782;  
 Best Local Similarity 32.4%; Pred. No. 5.3e-34;  
 Matches 118; Conservative 57; Mismatches 87; Indels 102; Gaps 12;

QY 23 PSLKSVQDQAYKAPVVVEGKV-----QG----- 45  
 ||: |||: | :| ||: ||| ||  
 Db 1 PSVGSVQELARRAAVVIEGKVHPPRRQQGALDRKAAGEAGAGARDQPVQDSPPSQDPLPA 60  
 QY 46 ---LVPAGGSSSNSTREPPASGRVALVKVLDKWPLRSGGLQREQVISV-----GS 92  
 :| || :: :| ||| | ::|||::: ::| |  
 Db 61 VNWTLP TGGPEPST--DQPGDPAPYLVKVHVQVAVKAGGLKKDSLTLTVRLDTWGHPAFFPS 118  
 QY 93 CVPLERNQRYIFFLEPT-----EQPLVFKTAFAPLDTNGKNLKEVGKILCTDCATRPKL 147  
 | |: : ||||: || | : :| ||: | :||| || :|| | :|  
 Db 119 CGRLKEDSRYIFFMEPDANSSGRAPPAFRASFPPLET-GRNLKKEVSRVLCKRCALPPRL 177  
 QY 148 KKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRS---RDIRIKYGNRKNRSL 204  
 |: ||| | | | :|| :: :|||: | ||| :||: | :| |

Db 178 KEMKSQESAAGSKLVLRCESTSEYSSLRFKWFKNGNELNRKNKPENIKIQKKPGK--SEL 235  
 QY 205 QFNKVKVEDAGEYVCEAENILGKDTVGRGL-----YV----- 236  
 : || : |:|:|:|: : || | : :  
 Db 236 RINKASLADSGEYMCKVISKLGNDASANITIVESNEFITGMPASTETAYVSSESPIRIS 295  
 QY 237 -----NSVSTTLSSWSG--HARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPNGFF 285  
 |: |:| :| :| | || | |:|:|:|:| |: :: :: | |||| |  
 Db 296 VSTEGANTSSSTSTSTTGTSHLIKCAEKEKTFCVNGGECFTVKDLSNPSRYLCKCPNEFT 355  
 QY 286 GQRC 289  
 | ||  
 Db 356 GDRC 359

# RESULT 2

## Q9ESA1

ID Q9ESA1 PRELIMINARY; PRT; 342 AA.  
 AC Q9ESA1;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Glial growth factor GGF beta 4 (Fragment).  
 GN NRG1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RC TISSUE=Axotomized lumbar dorsal root ganglion/spinal cord;  
 RA Carroll S.L., Stonecypher M.S., Anderson K.D., Pearson R.J. Jr.,  
 RA Frohnert P.W.;  
 RT "Structural and Functional Diversity of Glial Growth Factor Isoforms  
 RT Expressed in Regenerating Peripheral Nerve and Associated Neurons.";  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBDJ databases.  
 DR EMBL; AF194997; AAG28451.1; -.  
 DR HSSP; Q12784; 1HRE.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR Pfam; PF00008; EGF; 1.  
 DR Pfam; PF00047; ig; 1.  
 DR SMART; SM00181; EGF; 1.  
 DR SMART; SM00408; IGc2; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 KW EGF-like domain; Immunoglobulin domain.  
 FT NON\_TER 1 1  
 FT NON\_TER 342 342  
 SQ SEQUENCE 342 AA; 37836 MW; 8BE36FC836553124 CRC64;

Query Match 26.3%; Score 460; DB 11; Length 342;  
 Best Local Similarity 33.8%; Pred. No. 2.8e-32;  
 Matches 106; Conservative 54; Mismatches 92; Indels 62; Gaps 10;

QY 43 VQGLVPAGGSSSNSTREPPASGRVALVKVLDKWPLRSGGLQREQVISV-----GS 92  
 | :| || : : | ||| | :|||::: :| |  
 Db 5 VNWTLP TGGPEPST--DQPGDPAPYLKVHVQVAVKAGGLKKDSLTLTVRLDTWGHAPFPS 62

QY 93 CVPLERNQRYIFFLEPT-----EQPLVFKTAFAPLDTNGKNLKKEVGKILCTDCATRPKL 147  
 | | : : |||||:| | | : | ||: | :||| || :| | | :|  
 Db 63 CGRLKEDSRYIFFMEPDANSSGRAPPAFRASFPPLET-GRNLKKEVSRVLCKRCALPPRL 121

QY 148 KKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRS---RDIRIKYGNRKN SRL 204  
 | :||| | | | :|| : : :|||:| ||| :| : | : | |  
 Db 122 KEMKSQESAAGSKLVLCETSSSEYSSLRFKWFKNGNELNRKNKPENIKIQKKPGK--SEL 179

QY 205 QFNKVKVEDAGEYVCEAENILGKDTVGRGL-----YV----- 236  
 : || : | :|||:| : || | : : ||  
 Db 180 RINKASLADSGEYMCKVISKLGND SASANITIVESNEFITGMPASTETAYVSSSESPIRIS 239

QY 237 -----NSVSTTLSSWSG--HARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPNGFF 285  
 | : | : | : | | | | | : : ||| | : : : | | ||| |  
 Db 240 VSTEGANTSSSTSTSTTGTSHLIKCAEKEKTFVCVNGGECFTVKDLNPSRYLCKCPNEFT 299

QY 286 GQRCLEKLPLRLYM 299  
 | || : ||  
 Db 300 GDRCQNYVMASFYM 313

# RESULT 3

## Q9ESA2

ID Q9ESA2 PRELIMINARY; PRT; 323 AA.  
 AC Q9ESA2;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Glial growth factor GGF beta 3 (Fragment).  
 GN NRG1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RA Carroll S.L., Stonecypher M.S., Anderson K.D., Pearson R.J. Jr.,  
 RA Frohnert P.W.;  
 RT "Structural and Functional Diversity of Glial Growth Factor Isoforms  
 RT Expressed in Regenerating Peripheral Nerve and Associated Neurons."  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF194996; AAG28450.1; -.  
 DR HSSP; Q12784; 1HRE.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR Pfam; PF00008; EGF; 1.  
 DR Pfam; PF00047; ig; 1.  
 DR SMART; SM00181; EGF; 1.  
 DR SMART; SM00408; IGc2; 1.

DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 KW EGF-like domain; Immunoglobulin domain.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 323 AA; 35358 MW; C7DF153A939A80C8 CRC64;

Query Match 25.8%; Score 452; DB 11; Length 323;  
 Best Local Similarity 34.2%; Pred. No. 1.3e-31;  
 Matches 104; Conservative 52; Mismatches 86; Indels 62; Gaps 10;

Qy 43 VQGLVPAGGSSSNSTREPPASGRVALVKVLDKWPLRSGGLQREQVISV-----GS 92  
 | :| || :: :| |||| |::|||::: ::| |  
 Db 5 VNWTLPTGGPEPST--DQPGDPAPYLVKVHQVWAVKAGGLKKDSLTVRLDTWGHPAFPS 62  
 Qy 93 CVPLERNQRYIFFLEPT-----EQPLVFKTAFAPLDTNGKNLKKEVGKILCTDCATRPKL 147  
 | |: : ||||:| | |: :| ||:| |:||||| :|| || |:|  
 Db 63 CGRLKEDSRYIFFMEPDANSSGRAPPAFRASFPPLET-GRNLKKEVSRVLCRRCALPPRL 121  
 Qy 148 KKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRS---RDIRIKYGNRKNRSL 204  
 |:|||| | | |:| : : :|||:| |||| :||: | : | |  
 Db 122 KEMKSQESAAGSKLVLRCESTSEYSSLRFKWFKNELNRKNKPNIKIQKKPGK--SEL 179  
 Qy 205 QFNKVKVEDAGEYVCEAENILGKDTVGRGL-----YV----- 236  
 : || |:|||:| : || | : : ||  
 Db 180 RINKASPADSGEYMCKVISKLGNDASANITIVESNEFITGMPASTETAYVSSESPIRIS 239  
 Qy 237 -----NSVSTTLSSWSG--HARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPNGFF 285  
 |: |:| :| :| | || | |:||||| |: : : : | |||| |  
 Db 240 VSTEGANTSSSTSTSTTGTSHLIKCAEKEKTFVNGGECFTVKDLNSPSRYLCKCPNEFT 299  
 Qy 286 GQRC 289  
 | ||  
 Db 300 GDRC 303

#### RESULT 4

##### Q9ESA3

ID Q9ESA3 PRELIMINARY; PRT; 317 AA.  
 AC Q9ESA3;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Glial growth factor GGF beta 2 (Fragment).  
 GN NRGL.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RC TISSUE=Axotomized lumbar dorsal root ganglion/spinal cord;  
 RA Carroll S.L., Stonecypher M.S., Anderson K.D., Pearson R.J. Jr.,  
 RA Frohnert P.W.;  
 RT "Structural and Functional Diversity of Glial Growth Factor Isoforms  
 RT Expressed in Regenerating Peripheral Nerve and Associated Neurons."  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AF194995; AAG28449.1; -.  
 DR HSSP; Q12784; 1HRE.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR000886; ER\_target\_S.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR Pfam; PF00008; EGF; 1.  
 DR Pfam; PF00047; ig; 1.  
 DR SMART; SM00181; EGF; 1.  
 DR SMART; SM00408; IGc2; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS00014; ER\_TARGET; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 KW EGF-like domain; Immunoglobulin domain.  
 FT NON\_TER 1 1  
 FT NON\_TER 317 317  
 SQ SEQUENCE 317 AA; 34785 MW; 4487FA3E9CD876B9 CRC64;

Query Match 25.7%; Score 450; DB 11; Length 317;  
 Best Local Similarity 33.9%; Pred. No. 2e-31;  
 Matches 103; Conservative 54; Mismatches 85; Indels 62; Gaps 10;

QY 43 VQGLVPAGGSSSNSTREPPASGRVALVKVLDKWPLRSGGLQREQVISV-----GS 92  
 | :| || :: :| ||| | ::|||::: ::| |  
 Db 5 VNWTLPTGGPEPST--DQPGDPAPYLKVHVQVWAVKAGGLKKDSLTLTVRLDTWGHPAFPS 62  
 QY 93 CVPLERNQRYIFFLEPT-----EQPLVFKTAFAPLDTNGKNLKEVGKILCTDCATRPKL 147  
 | |: : ||||:| | |: :| ||:| |:||||| ::|| || |:|  
 Db 63 CGRLKEDSRYIFFMEPDANSSGRAPPAFRASFPLET-GRDLKKEVSRVLCKRCALPPRL 121  
 QY 148 KKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRS---RDIRIKYGNRKN SRL 204  
 |:||| | | |:|| :: ::|||:| ||| |:|: |: | |  
 Db 122 KEMKSQESAAGSKLVLRCE TSSEYSSLRFKWFKNGNELNRKNKPENIKIQKKPGK--SEL 179  
 QY 205 QFNKVKVEDAGEYVCEAENILGKDTVGRGL-----YV----- 236  
 : || : |:| |:|: : || |: : ||  
 Db 180 RINKASLADSGEYMCKVISKLGND SASANITIVESNEFITGMPASTETAYVSSESPIRIS 239  
 QY 237 -----NSVSTTLSSWSG--HARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPNGFF 285  
 |: |:| :| :| | || | |:| |||| |: :: :: | |||| |  
 Db 240 VSTEGANTSSSTSTSTTGTSHLIKCAEKEKTF CVNGGECFTVKDLSNPSRYLCKCPNEFT 299  
 QY 286 GQRC 289  
 | ||  
 Db 300 GDRC 303

# RESULT 5 Q810X2

ID Q810X2 PRELIMINARY; PRT; 79 AA.  
 AC Q810X2;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Neuregulin 2-alpha (Fragment).  
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CD-1; TISSUE=Olfactory bulb;  
 RA Mautino B., Dalla Costa L., Dati C.;  
 RT "Bioactive recombinant NRG1, NRG2 and NRG3 expressed in E. coli."  
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY227025; AAO72522.1; -.  
 DR InterPro; IPR006209; EGF\_like.  
 DR Pfam; PF00008; EGF; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 79 AA; 8727 MW; DA4501900C610780 CRC64;

Query Match 25.4%; Score 444; DB 11; Length 79;  
 Best Local Similarity 100.0%; Pred. No. 1e-31;  
 Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 KCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQSVLWDTP 311  
 Db 1 KCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQSVLWDTP 60  
 QY 312 GTGVSSSQWSTSPSTLDLN 330  
 Db 61 GTGVSSSQWSTSPSTLDLN 79

# RESULT 6

Q8NFN3

ID Q8NFN3 PRELIMINARY; PRT; 348 AA.  
 AC Q8NFN3;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Neuregulin 1 isoform GGF2 (Fragment).  
 GN NRG1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Stefansson H., Sigurdsson E., Steinthorsdottir V., Bjornsdottir S.,  
 RA Sigmundsson T., Ghosh S., Brynjolfsson J., Gunnarsdottir S.,  
 RA Ivarsson O., Chou T.T., Hjaltason O., Birgisdottir B., Jonsson H.,  
 RA Gudnadottir V.G., Gudmundsdottir E., Bjornsson A., Ingvarsson B.,  
 RA Ingason A., Sigfusson S., Hardardottir H., Harvey R.P., Brunner D.,  
 RA Mutel V., Gonzalo A., Lemke G., Sainz J., Johannesson G.,  
 RA Andresson T., Gudbjartsson D., Manolescu A., Frigge M.L., Gurney M.E.,  
 RA Kong A., Gulcher J.R., Petursson H., Stefansson K.;  
 RT "Neuregulin 1 and susceptibility to Schizophrenia."  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF491780; AAM71140.1; -.  
 DR InterPro; IPR003599; Ig.





RL Mol. Carcinog. 21:156-163(1998).  
 CC -!- FUNCTION: DIRECT LIGAND FOR ERBB3 AND ERBB4 TYROSINE KINASE  
 CC RECEPTORS. CONCOMITANTLY RECRUITS ERBB1 AND ERBB2 CORECEPTORS,  
 CC RESULTING IN LIGAND-STIMULATED TYROSINE PHOSPHORYLATION AND  
 CC ACTIVATION OF THE ERBB RECEPTORS. MAY PLAY AN IMPORTANT ROLE IN  
 CC PROVIDING GROWTH ADVANTAGE IN NEOPLASTIC CELLS.  
 CC -!- SUBUNIT: THE CYTOPLASMIC DOMAIN INTERACTS WITH THE LIM DOMAIN  
 CC REGION OF LIMK1 (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: EXISTS AS TYPE I MEMBRANE PROTEIN AND AS A  
 CC PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE MEMBRANE-  
 CC BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).  
 CC -!- TISSUE SPECIFICITY: EXPRESSED AT HIGHER LEVEL AFTER NEOPLASMIC  
 CC TRANSFORMATION OF CELLS.  
 CC -!- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION  
 CC OF TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE  
 CC PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN  
 CC DIMERIZATION (BY SIMILARITY).  
 CC -!- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE  
 CC DOMAIN.  
 CC -!- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE PLASMA MEMBRANE ON THE  
 CC EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR  
 CC FORM (BY SIMILARITY).  
 CC -!- PTM: EXTENSIVE GLYCOSYLATION PRECEDES PROTEOLYTIC CLEAVAGE (BY  
 CC SIMILARITY).  
 CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.  
 CC -!- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.  
 DR EMBL; U96612; AAB71812.1; -.  
 DR HSSP; Q12784; 1HRE.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0008083; F:growth factor activity; IEA.  
 DR GO; GO:0009790; P:embryonic development; IEA.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR InterPro; IPR002154; Neuregulin.  
 DR Pfam; PF00008; EGF; 1.  
 DR Pfam; PF00047; ig; 1.  
 DR Pfam; PF02158; Neuregulin; 1.  
 DR PRINTS; PR01089; NEUREGULIN.  
 DR SMART; SM00181; EGF; 1.  
 DR SMART; SM00408; IGc2; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 KW Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;  
 KW Transmembrane; Alternative splicing.  
 FT PROPEP 1 13 BY SIMILARITY.  
 FT CHAIN 14 461 PRO-NEUREGULIN-1, MEMBRANE-BOUND FORM.  
 FT CHAIN 14 241 NEUREGULIN-1.  
 FT DOMAIN 14 242 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 243 265 INTERNAL SIGNAL SEQUENCE (POTENTIAL).  
 FT DOMAIN 266 461 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 50 119 IG-LIKE C2-TYPE DOMAIN.  
 FT DOMAIN 165 177 SER/THR-RICH.  
 FT DOMAIN 178 222 EGF-LIKE.

FT	DISULFID	57	112	BY SIMILARITY.
FT	DISULFID	182	196	BY SIMILARITY.
FT	DISULFID	190	210	BY SIMILARITY.
FT	DISULFID	212	221	BY SIMILARITY.
FT	CARBOHYD	73	73	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	120	120	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	126	126	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	164	164	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	461 AA; 50890 MW; 935C9560F7148336 CRC64;		

Query Match 16.6%; Score 290; DB 11; Length 461;  
 Best Local Similarity 33.2%; Pred. No. 4.6e-17;  
 Matches 65; Conservative 33; Mismatches 58; Indels 40; Gaps 5;

Qy	142	ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELN-RSRDIRIKYGNGRK	200
		: : :        : :  :  : : :        :	
Db	34	ALPPRLKEMKIQESAAGSKLVLRCESTSSEYPELRFKWFKNKNGSELNKRTPQNIKLQKKPG	93
Qy	201	NSRLQFNKVKVEDAGEYVCEAENILGKDTVIRG-----RLYV-----	236
		:    : : : :  :      :	
Db	94	KSELRINKASLADSGEYMCKVISKLGNDASANITIVDSNEFITGMPASTERAYVSSESP	153
Qy	237	-----NSVSTTLSSWSG--HARKCNETAKSYCVNGGVYIEGINQLS---CKCP	281
		: :  :  :        : : :    :  :  :	
Db	154	IRISVSTEGANTSSSTSTTTGTSHLVKCAEKEKTFVCVNGGECFMVKDLSNPSRYLCKCQ	213
Qy	282	NGFFGQRCLEKLPLRL	297
		: : :	
Db	214	PGFTGARCTENVPMKV	229

# RESULT 8

Q07112

ID Q07112 PRELIMINARY; PRT; 241 AA.  
 AC Q07112;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Glial growth factor.  
 GN GGFBBP5.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Posterior pituitary;  
 RX MEDLINE=93205115; PubMed=8096067;  
 RA Marchionni M.A., Goodearl A.D.G., Chen M., Bermingham-McDonogh O.,  
 RA Kirk C., Hendricks M., Danehy F., Misumi D., Sudhalter J.,  
 RA Kobayashi K., Wroblewski D., Lynch C., Baldassarre M., Hiles I.,  
 RA Davis J.B., Hsuan J., Totty N.F., Otsu M., McBurney R.N.,  
 RA Waterfield M.D., Stroobant P., Gwynne D.;  
 RT "Glial growth factors are alternatively spliced erbB2 ligands  
 RT expressed in the nervous system."  
 RL Nature 362:312-318(1993).



RL Nature 420:563-573(2002).  
 DR EMBL; AK051824; BAC34784.1; -.  
 DR MGD; MGI:96083; Nrg1.  
 DR GO; GO:0005737; C:cytoplasm; IDA.  
 DR GO; GO:0005887; C:integral to plasma membrane; IDA.  
 DR GO; GO:0045202; C:synaptic junction; IDA.  
 DR GO; GO:0005176; F:Neu/ErbB-2 receptor binding; IDA.  
 DR GO; GO:0016477; P:cell migration; IGI.  
 DR GO; GO:0000902; P:cellular morphogenesis; IDA.  
 DR GO; GO:0010001; P:glial cell differentiation; IMP.  
 DR GO; GO:0007507; P:heart development; IDA.  
 DR GO; GO:0007626; P:locomotory behavior; IMP.  
 DR GO; GO:0000165; P:MAPKKK cascade; IDA.  
 DR GO; GO:0007517; P:muscle development; IMP.  
 DR GO; GO:0042055; P:neuronal lineage restriction; IMP.  
 DR GO; GO:0045213; P:neurotransmitter receptor metabolism; IMP.  
 DR GO; GO:0007422; P:peripheral nervous system development; IMP.  
 DR GO; GO:0045860; P:positive regulation of protein kinase activity; IDA.  
 DR GO; GO:0046579; P:positive regulation of RAS protein signal t. . .; IDA.  
 DR GO; GO:0045595; P:regulation of cell differentiation; IMP.  
 DR GO; GO:0007416; P:synaptogenesis; IMP.  
 DR InterPro; IPR003599; Ig.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR Pfam; PF00047; ig; 1.  
 DR SMART; SM00409; IG; 1.  
 DR SMART; SM00408; IGc2; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 SQ SEQUENCE 211 AA; 22893 MW; 75D3674B988BE0D3 CRC64;

Query Match 13.4%; Score 234; DB 11; Length 211;  
 Best Local Similarity 31.7%; Pred. No. 1.4e-12;  
 Matches 57; Conservative 32; Mismatches 47; Indels 44; Gaps 6;

QY 142 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNR---SRDIRIKYNGG 198  
 | | :||:|||| | | :|| :: :|||:| |||| ::||: |  
 Db 34 ALPPRLKEMKSEQSAAGSKLVLCETSSEYSSLRFKWFKNGNELNRRNKPQNVKIQKKPG 93  
 QY 199 RKNSRLQFNKVKVEDAGEYVCEAENILGKDTVGR-----RLYV--- 236  
 : | |: || : |:|||:|: : || |: | ||  
 Db 94 K--SELRINKASLADSGEYMCKVISKLGNDSASANITIVESNDLTTGMSASTERPYVSSE 151  
 QY 237 -----NSVSTTLSSWSG--HARKCNETAKSYCVNGGVCYYIEGINQLS---CK 279  
 |: |:| :| :| | || | |:| |||| |: :: :: | ||  
 Db 152 SPIRISVSTEGANTSSSTSTSTGTSHLIKCAEKEKTFVNGGECFMVKDLSNPSRYLCK 211

# RESULT 10

Q9ESA4

ID Q9ESA4 PRELIMINARY; PRT; 244 AA.

AC Q9ESA4;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)

DE Glial growth factor (Fragment).

GN NRGL.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RA Carroll S.L., Stonecypher M.S., Anderson K.D., Pearson R.J. Jr.,  
 RA Frohnert P.W.;  
 RT "Structural and Functional Diversity of Glial Growth Factor Isoforms  
 RT Expressed in Regenerating Peripheral Nerve and Associated Neurons."  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF194994; AAG28434.1; -.  
 FT NON\_TER 244 244  
 SQ SEQUENCE 244 AA; 25866 MW; 019CBC2DFFF8F625 CRC64;

Query Match 12.6%; Score 221; DB 11; Length 244;  
 Best Local Similarity 29.5%; Pred. No. 2.5e-11;  
 Matches 62; Conservative 28; Mismatches 44; Indels 76; Gaps 8;

QY 5 PAPGFSMLLFGVSL-----ACYS--PSLKSVDQAYKAPVVVEGKV--- 43  
 | | :| | : || | | :| | | :| | :| | |  
 Db 38 PPPLLLLLLLGTAALAPGAAAERAAPAGASVCYSSPPSVGSGVQELARRAAVVIEGKVHPP 97  
 QY 44 ---QG-----LVPAGGSSSNSTREPPASGRV 66  
 || :| || :: :|  
 Db 98 RRQQGALDRKAAGEAGAGARDQPVQDSPPSQDPLPAVNWTLPTGGPEPST--DQPGDPAP 155  
 QY 67 ALVKVLDKWPLRSGGLQREQVISV-----GSCVPLERNQRYIFFLEPT-----EQ 111  
 ||| | ::|||:: ::| || | : : ||| :||  
 Db 156 YLVKVHVQVWAVKAGGLKKDSLTLVRLDTWGHAPAFPSCGRLKEDSRYIFFMEPDANSSGRA 215  
 QY 112 PLVFKTAFAPLDTNGKNLKKEVGKILCTDC 141  
 | | : :| ||:| |:||||| :|| |  
 Db 216 PPAFRASFPPLET-GRNLKKEVSRVLCKRC 244

# RESULT 11

## Q810X1

ID Q810X1 PRELIMINARY; PRT; 54 AA.  
 AC Q810X1;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Neuregulin 2-beta (Fragment).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CD-1; TISSUE=Olfactory bulb;  
 RA Mautino B., Dalla Costa L., Dati C.;  
 RT "Bioactive recombinant NRG1, NRG2 and NRG3 expressed in E. coli."  
 RL Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AY227026; AAO72523.1; -.  
 DR InterPro; IPR006209; EGF\_like.  
 DR Pfam; PF00008; EGF; 1.

DR PROSITE; PS00022; EGF\_1; 1.  
DR PROSITE; PS01186; EGF\_2; 1.  
FT NON\_TER 1 1  
FT NON\_TER 54 54  
SQ SEQUENCE 54 AA; 6019 MW; C25AA17A4D0BA59A CRC64;

Query Match 11.4%; Score 200; DB 11; Length 54;  
Best Local Similarity 89.5%; Pred. No. 2.3e-10;  
Matches 34; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 252 KCNETAKSYCVNGGVVCYYIEGINQLSCKCPNGFFGQRC 289  
||||| : |||  
Db 1 KCNETAKSYCVNGGVVCYYIEGINQLSCKCPVGYTGDRC 38

# RESULT 12

## Q8NFN2

ID Q8NFN2 PRELIMINARY; PRT; 167 AA.  
AC Q8NFN2;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Neuregulin 1 isoform GGF (Fragment).  
GN NRG1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Stefansson H., Sigurdsson E., Steinthorsdottir V., Bjornsdottir S.,  
RA Sigmundsson T., Ghosh S., Brynjolfsson J., Gunnarsdottir S.,  
RA Ivarsson O., Chou T.T., Hjaltason O., Birgisdottir B., Jonsson H.,  
RA Gudnadottir V.G., Gudmundsdottir E., Bjornsson A., Ingvarsson B.,  
RA Ingason A., Sigfusson S., Hardardottir H., Harvey R.P., Brunner D.,  
RA Mutel V., Gonzalo A., Lemke G., Sainz J., Johannesson G.,  
RA Andresson T., Gudbjartsson D., Manolescu A., Frigge M.L., Gurney M.E.,  
RA Kong A., Gulcher J.R., Petursson H., Stefansson K.;  
RT "Neuregulin 1 and susceptibility to Schizophrenia.";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF491780; AAM71139.1; -.  
DR InterPro; IPR003599; Ig.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003598; Ig\_c2.  
DR Pfam; PF00047; ig; 1.  
DR SMART; SM00409; IG; 1.  
DR SMART; SM00408; IGc2; 1.  
DR PROSITE; PS50835; IG\_LIKE; 1.  
KW Immunoglobulin domain.  
FT NON\_TER 167 167  
SQ SEQUENCE 167 AA; 17983 MW; 9C2FB3A579325FF4 CRC64;

Query Match 10.3%; Score 180.5; DB 4; Length 167;  
Best Local Similarity 30.9%; Pred. No. 5.6e-08;  
Matches 51; Conservative 30; Mismatches 63; Indels 21; Gaps 5;

QY 126 GKNLKKEVGKILCTDCAT-----RPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRW 178

```

Db      11 GKGKKKERGSGKKPESAAGSQSPALPPRLKEMKSQESAAGSKLVLRCEISSEYSSLRFKW 70
Qy      179 FKDGKELNRS---RDIRIKYGNRKNRSLQFNKVVEDAGEYVCEAENILGKDTVGRGLY 235
Db      71 FKNGNELNRKNKPQNIKIQKKPGK--SELRINKASLADSGEYMCKVISKLGNDSSASANIT 128
Qy      236 V---NSVSTTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLS 277
Db      129 IVESNEIITGMPASTEGAYVSSESPIRISVS-----TEGANTSS 167

```

RESULT 13

Q86GD6

```

ID      Q86GD6          PRELIMINARY;          PRT; 8625 AA.
AC      Q86GD6;
DT      01-JUN-2003 (TrEMBLrel. 24, Created)
DT      01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Projectin.
GN      PROJ.
OS      Procambarus clarkii (Red swamp crayfish).
OC      Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC      Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC      Astacoidea; Cambaridae; Procambarus.
OX      NCBI_TaxID=6728;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Muscle;
RA      Oshino T., Shimamura J., Fukuzawa A., Maruyama K., Kimura S.;
RT      "The entire cDNA sequences of projectin isoforms of crayfish claw
RT      closer and flexor muscles and their localization.";
RL      J. Muscle Res. Cell. Motil. 0:0-0(2003).
DR      EMBL; AB055927; BAC66140.1; -.
DR      GO; GO:0005743; C:mitochondrial inner membrane; IEA.
DR      GO; GO:0005524; F:ATP binding; IEA.
DR      GO; GO:0005488; F:binding; IEA.
DR      GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR      GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR      GO; GO:0003743; F:translation initiation factor activity; IEA.
DR      GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR      GO; GO:0006413; P:translational initiation; IEA.
DR      GO; GO:0006810; P:transport; IEA.
DR      InterPro; IPR003962; FnIII_subd.
DR      InterPro; IPR003961; FN_III.
DR      InterPro; IPR008957; FN_III-like.
DR      InterPro; IPR003599; Ig.
DR      InterPro; IPR007110; Ig-like.
DR      InterPro; IPR003598; Ig_c2.
DR      InterPro; IPR003596; Ig_v.
DR      InterPro; IPR001993; Mitoch_carrier.
DR      InterPro; IPR000719; Prot_kinase.
DR      InterPro; IPR002290; Ser_thr_pkinase.
DR      InterPro; IPR008271; Ser_thr_pkin_AS.
DR      InterPro; IPR001950; TIF_SUI1.
DR      InterPro; IPR001245; Tyr_pkinase.
DR      Pfam; PF00041; fn3; 39.

```

DR Pfam; PF00047; ig; 13.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00014; FNTYPEIII.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR SMART; SM00060; FN3; 39.  
 DR SMART; SM00409; IG; 36.  
 DR SMART; SM00408; IGc2; 24.  
 DR SMART; SM00406; IGv; 3.  
 DR SMART; SM00220; S\_TKc; 1.  
 DR SMART; SM00219; TyrKc; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 24.  
 DR PROSITE; PS00215; MITOCH\_CARRIER; 2.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE; PS01118; SUI1\_1; 1.  
 SQ SEQUENCE 8625 AA; 962637 MW; 56B8E4C4FE0AFC90 CRC64;

Query Match 9.5%; Score 166; DB 5; Length 8625;  
 Best Local Similarity 24.2%; Pred. No. 0.00022;  
 Matches 87; Conservative 39; Mismatches 130; Indels 104; Gaps 16;

Qy 2 RRDP-----APGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPA 49  
 || | || | : | : | | : || ||  
 Db 8115 RRQPLTYKLWQEEGEGAPSFTFLLRPRVIQCH-----QTCKLLCCLAGKP---VPT 8162  
 Qy 50 ----GGSSSNST---REPPASGRVALVKVLDKWPLRSGGLQREQVISVG----SCVPLER 98  
 || | : || | : : : | || : | : || | :  
 Db 8163 VKWYKGSQELSKFDYSQSHADG-VVTIEIVNCKPADSGKYRCVATNSLGTDETSCVVIVE 8221  
 Qy 99 NQRYIFF----LEPTEQPLV-----FKTAFAPLDTNGK---NLKKEVGKILCTDCATRP 145  
 : || | | | : : : : | : : | | | | |  
 Db 8222 DRRYIETTIKDLPPPPTPAIRVDDTSSSYFTSTHKDGRSSTSTKVEAASSSSTSSAAAS 8281  
 Qy 146 KLK-----KMKSQTGQV-----GEKQS 162  
 | : || ||  
 Db 8282 GAKRTLKPYGKRQDSTGSTSRSRSATKELELPPDDSLMGPPGFSGELPKTLAIKDGEALC 8341  
 Qy 163 LKCEAAAGNPQPSYRWFKD GKELNRSRDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAE 222  
 || | : : | | | | : | | : | | : | | | : |  
 Db 8342 LKC-TVKGDPEPQVSWFKDGEPLSSSDIIDLKYRQGL--ASLTINEVFPEDGLYVCKAT 8398  
 Qy 223 NILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAK---SYCVNGGVCYYIEGINQLSCK 279  
 : || : : : : : || | : | | : ||||  
 Db 8399 SSLGSAETKCKLSISPMEQQINGKSGRGDKLPRITQHLLSQEVPDGTAH-----TL SCK 8452

# RESULT 14

Q8IOL3

ID Q8IOL3 PRELIMINARY; PRT; 5175 AA.

AC Q8IOL3;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE C. elegans him-4 protein (corresponding sequence F15G9.4a).

GN F15G9.4 OR HIM-4.

OS Caenorhabditis elegans.



OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sulston J.E.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology.";  
 RL Science 282:2012-2018(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Kershaw J.K.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; Z47068; CAA87335.1; -.  
 DR EMBL; Z47070; CAA87335.1; JOINED.  
 DR EMBL; Z47070; CAA87344.1; -.  
 DR EMBL; Z47068; CAA87344.1; JOINED.  
 DR PIR; T20992; T20992.  
 DR WormPep; F15G9.4a; CE18595.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005509; F:calcium ion binding; IEA.  
 DR GO; GO:0005215; F:transporter activity; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR000152; Asx\_hydroxyl\_S.  
 DR InterPro; IPR000515; BPD\_transp.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR003599; Ig.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR InterPro; IPR003596; Ig\_v.  
 DR Pfam; PF00047; ig; 47.  
 DR SMART; SM00181; EGF; 2.  
 DR SMART; SM00179; EGF\_CA; 2.  
 DR SMART; SM00409; IG; 45.  
 DR SMART; SM00408; IGc2; 47.  
 DR SMART; SM00406; IGv; 12.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE; PS00402; BPD\_TRANSP\_INN\_MEMBR; 1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01187; EGF\_CA; 2.  
 DR PROSITE; PS50835; IG\_LIKE; 47.  
 SQ SEQUENCE 5175 AA; 568471 MW; 4B2561803BBC62A4 CRC64;

Query Match 9.4%; Score 164.5; DB 5; Length 5175;  
 Best Local Similarity 26.5%; Pred. No. 0.00015;  
 Matches 62; Conservative 26; Mismatches 79; Indels 67; Gaps 10;

QY 49 AGGSSSNSTR----EPPASGRVALVKVLDK-----WPLRS 79  
 ||| |: | ||| : ||| : | :  
 Db 595 AGGMSTRKMRDLIMEPPS-----VKVTPQDVYFNMREGVNLSCEAMGDPKPEVHWYFKG 648

QY 80 GGLQREQVISVGSCVPLERNQRYIFFL-----EPTEQPLVFKTAFAPLDTNGKNLKKEV 133  
 | : || :: :::: | | : | | |  
 Db 649 RHLLNDYKYQVG-----QDSKFLYIRDATHHDEGTYECRAMSQAGQARDTTDLML---- 698  
 QY 134 GKILCTDCATRPKLK--KMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDI 191  
 || ||:: : | | : |:: | :| : | | | |||:|:| : |  
 Db 699 -----ATPPKVEIIQNKMVGR-GDRVSFECKTIRGKPHPKIRWFKNGKDLIKPDDY 749  
 QY 192 RIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSS 245  
 || | :| | ||| | | ||: || | | | : |  
 Db 750 -IKINEG-----QLHIMGAKDEDAGAYSCVGENMAGKDVQVANLSVGRVPTIIES 798

RESULT 15

O76518

ID O76518 PRELIMINARY; PRT; 5198 AA.  
 AC O76518; Q10036;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hemicentin precursor (C. elegans him-4 protein) (corresponding  
 DE sequence F15G9.4b).  
 GN F15G9.4 OR HIM-4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX PubMed=11222143;  
 RA Vogel B.E., Hedgecock E.M.;  
 RT "Hemicentin, a conserved extracellular member of the immunoglobulin  
 RT superfamily, organizes epithelial and other cell attachments into  
 RT oriented line-shaped junctions.";  
 RL Development 128:883-894(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Sulston J.E.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology.";  
 RL Science 282:2012-2018(1998).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Kershaw J.K.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF074901; AAC26792.1; -.  
 DR EMBL; Z47068; CAA87336.1; -.  
 DR EMBL; Z47070; CAA87336.1; JOINED.  
 DR EMBL; Z47070; CAA87345.1; -.  
 DR EMBL; Z47068; CAA87345.1; JOINED.

DR PIR; T43290; T43290.  
 DR HSSP; P00736; 1APQ.  
 DR WormPep; F15G9.4b; CE18596.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005509; F:calcium ion binding; IEA.  
 DR GO; GO:0005215; F:transporter activity; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR000152; Asx\_hydroxyl\_S.  
 DR InterPro; IPR000515; BPD\_transp.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF00047; ig; 47.  
 DR SMART; SM00179; EGF\_CA; 1.  
 DR SMART; SM00408; IGc2; 44.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE; PS00402; BPD\_TRANSP\_INN\_MEMBR; 1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01187; EGF\_CA; 2.  
 DR PROSITE; PS50835; IG\_LIKE; 47.  
 KW EGF-like domain; Immunoglobulin domain; Signal.  
 FT SIGNAL 1 24 POTENTIAL.  
 FT CHAIN 25 5198 HEMICENTIN.  
 SQ SEQUENCE 5198 AA; 570809 MW; DA8511FF2B58D37B CRC64;

Query Match 9.4%; Score 164.5; DB 5; Length 5198;  
 Best Local Similarity 26.5%; Pred. No. 0.00015;  
 Matches 62; Conservative 26; Mismatches 79; Indels 67; Gaps 10;

QY 49 AGGSSSNSTR---EPPASGRVALVKVLDK-----WPLRS 79  
 ||| |: | ||: ||| :  
 Db 595 AGGMSTRKMRLDIMEPPS-----VKVTPQDVYFNMREGVNLSCEAMGDPKPEVHWYFKG 648  
 QY 80 GGLQREQVISVGSCVPLERNQRYIFFL-----EPTEQPLVFKTAFAPLDTNGKNLKKEV 133  
 | : || :: :: | | : | | |  
 Db 649 RHLNDYKYQVG-----QDSKFLYIRDATHHDEGTYECRAMSQAGQARDTTDLML----- 698  
 QY 134 GKILCTDCATRPKLK--KMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDI 191  
 || ||:: : | |: |:: | :: | | | |||::|::| : |  
 Db 699 -----ATPPKVEIIQNKMVGR-GDRVSFECKTIRGKPHPKIRWFKNGKDLIKPDY 749  
 QY 192 RIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSS 245  
 || | : | |||| | | ||: || | | | : |  
 Db 750 -IKINEG----QLHIMGAKDEDAGAYSCVGENMAGKDVQVANLSVGRVPTIIES 798

Search completed: August 17, 2004, 14:12:40  
 Job time : 39.3089 secs

OM protein - protein search, using sw model

Run on: August 17, 2004, 13:56:40 ; Search time 9.4586 Seconds  
(without alignments)  
1816.670 Million cell updates/sec

Title: US-09-864-675-2  
Perfect score: 1749  
Sequence: 1 MRRDPAPGFSMLLFGVSLAC.....PGTGVSSSQWSTSPSTLDLN 330

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1610	92.1	850	1	NRG2_HUMAN	O14511 homo sapien
2	1576	90.1	868	1	NRG2_RAT	O35569 rattus norv
3	1475	84.3	756	1	NRG2_MOUSE	P56974 mus musculu
4	311.5	17.8	677	1	NRG1_XENLA	O93383 xenopus lae
5	293.5	16.8	639	1	NRG1_HUMAN	Q02297 h pro-neure
6	284.5	16.3	602	1	NRG1_CHICK	Q05199 gallus gall
7	283	16.2	662	1	NRG1_RAT	P43322 r pro-neure
8	216.5	12.4	623	1	VEIN_DROME	Q94918 drosophila
9	155	8.9	298	1	JAM2_HUMAN	P57087 homo sapien
10	145	8.3	338	1	LAMP_RAT	Q62813 rattus norv
11	144	8.2	338	1	LAMP_HUMAN	Q13449 homo sapien
12	141.5	8.1	353	1	CEPU_CHICK	Q90773 gallus gall
13	140	8.0	761	1	NCA2_HUMAN	P13592 homo sapien
14	140	8.0	848	1	NCA1_HUMAN	P13591 homo sapien
15	139.5	8.0	6632	1	UN89_CAEEL	O01761 caenorhabdi
16	139	7.9	2012	1	DSCA_HUMAN	O60469 homo sapien
17	135.5	7.7	1356	1	VGR2_HUMAN	P35968 homo sapien

18	135	7.7	4391	1	PGBM_HUMAN	P98160	homo sapien
19	133.5	7.6	853	1	NCA1_BOVIN	P31836	bos taurus
20	133.5	7.6	1367	1	VGR2_MOUSE	P35918	mus musculu
21	131.5	7.5	338	1	LAMP_CHICK	Q98919	gallus gall
22	131.5	7.5	1343	1	VGR2_RAT	O08775	rattus norv
23	131	7.5	296	1	SMDF_HUMAN	Q15491	homo sapien
24	131	7.5	345	1	OPCM_RAT	P32736	rattus norv
25	131	7.5	824	1	MLT1_HUMAN	Q9udy8	homo sapien
26	130.5	7.5	837	1	NCM2_MOUSE	O35136	mus musculu
27	130.5	7.5	1040	1	AXO1_RAT	P22063	rattus norv
28	129.5	7.4	1018	1	CONT_HUMAN	Q12860	homo sapien
29	129.5	7.4	1091	1	NCA1_CHICK	P13590	gallus gall
30	129	7.4	345	1	OPCM_HUMAN	Q14982	homo sapien
31	129	7.4	1036	1	AXO1_CHICK	P28685	gallus gall
32	128.5	7.3	1010	1	CONT_CHICK	P14781	gallus gall
33	128.5	7.3	1020	1	CONT_MOUSE	P12960	mus musculu
34	128.5	7.3	1021	1	CONT_RAT	Q63198	rattus norv
35	128	7.3	1040	1	AXO1_HUMAN	Q02246	homo sapien
36	127	7.3	298	1	JAM1_BOVIN	Q9xt56	bos taurus
37	127	7.3	344	1	NTRI_RAT	Q62718	rattus norv
38	126	7.2	1217	1	EGF_MOUSE	P01132	mus musculu
39	124	7.1	344	1	NTRI_MOUSE	Q99pj0	mus musculu
40	123.5	7.1	858	1	NCA1_RAT	P13596	rattus norv
41	123	7.0	345	1	OPCM_BOVIN	P11834	bos taurus
42	122	7.0	344	1	NTRI_HUMAN	Q9p121	homo sapien
43	121.5	6.9	725	1	NCA2_MOUSE	P13594	mus musculu
44	121.5	6.9	837	1	NCM2_HUMAN	O15394	homo sapien
45	121.5	6.9	1115	1	NCA1_MOUSE	P13595	mus musculu

# ALIGNMENTS

## RESULT 1

### NRG2\_HUMAN

ID NRG2\_HUMAN STANDARD; PRT; 850 AA.  
AC O14511;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Pro-neuregulin-2 precursor (Pro-NRG2) [Contains: Neuregulin-2 (NRG-2)  
DE (Neural-and thymus-derived activator for ERBB kinases) (NTAK)  
DE (Divergent of neuregulin 1) (DON-1)].  
GN NRG2 OR NTAK.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Neuroblastoma;  
RX MEDLINE=98006324; PubMed=9348101;  
RA Higashiyama S., Horikawa M., Yamada K., Ichino N., Nakano N.,  
RA Nakagawa T., Miyagawa J., Matsushita N., Nagatsu T., Taniguchi N.,  
RA Ishiguro H.;  
RT "A novel brain-derived member of the epidermal growth factor family  
RT that interacts with ErbB3 and ErbB4.";

RL J. Biochem. 122:675-680(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS DON-1B AND DON-1R).  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=97342638; PubMed=9199335;  
 RA Busfield S.J., Michnick D.A., Chickering T.W., Revett T.L., Ma J.,  
 RA Woolf E.A., Comrack C.A., Dussault B.J., Woolf J., Goodearl A.D.J.,  
 RA Gearing D.P.;  
 RT "Characterization of a neuregulin-related gene, Don-1, that is highly  
 RT expressed in restricted regions of the cerebellum and hippocampus.";  
 RL Mol. Cell. Biol. 17:4007-4014(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4; 5 AND 6).  
 RC TISSUE=Fetal brain, and Lung;  
 RX MEDLINE=99295836; PubMed=10369162;  
 RA Ring H.Z., Chang H., Guilbot A., Brice A., LeGuern E., Francke U.;  
 RT "The human neuregulin 2 (NRG2) gene: cloning, mapping and evaluation  
 RT as a candidate for the autosomal recessive form of Charcot-Marie-Tooth  
 RT disease linked to 5q.";  
 RL Hum. Genet. 104:326-332(1999).  
 CC -!- FUNCTION: Direct ligand for ERBB3 and ERBB4 tyrosine kinase  
 CC receptors. Concomitantly recruits ERBB1 and ERBB2 coreceptors,  
 CC resulting in ligand-stimulated tyrosine phosphorylation and  
 CC activation of the ERBB receptors. May also promote the  
 CC heterodimerization with the EGF receptor.  
 CC -!- SUBCELLULAR LOCATION: EXISTS AS AN TYPE I MEMBRANE PROTEIN AND AS  
 CC A PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE  
 CC MEMBRANE-BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=8;  
 CC Name=1;  
 CC IsoId=O14511-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=O14511-2; Sequence=VSP\_003453;  
 CC Name=3;  
 CC IsoId=O14511-3; Sequence=VSP\_003455;  
 CC Name=4;  
 CC IsoId=O14511-4; Sequence=VSP\_003454;  
 CC Name=5;  
 CC IsoId=O14511-5; Sequence=VSP\_003458, VSP\_003459;  
 CC Name=6;  
 CC IsoId=O14511-6; Sequence=VSP\_003456, VSP\_003457;  
 CC Name=DON-1B;  
 CC IsoId=O14511-7; Sequence=VSP\_003452, VSP\_003455;  
 CC Name=DON-1R;  
 CC IsoId=O14511-8; Sequence=VSP\_003451;  
 CC -!- TISSUE SPECIFICITY: Restricted to the cerebellum in the adult.  
 CC -!- DOMAIN: The cytoplasmic domain may be involved in the regulation  
 CC of trafficking and proteolytic processing. Regulation of the  
 CC proteolytic processing involves initial intracellular domain  
 CC dimerization (By similarity).  
 CC -!- DOMAIN: ERBB receptor binding is elicited entirely by the EGF-like  
 CC domain (By similarity).  
 CC -!- PTM: Proteolytic cleavage close to the plasma membrane on the  
 CC external face leads to the release of the soluble growth factor  
 CC form (By similarity).  
 CC -!- PTM: Extensive glycosylation precedes the proteolytic cleavage (By

CC similarity).  
 CC -!- SIMILARITY: Contains 1 EGF-like domain.  
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.  
 CC -!- SIMILARITY: Belongs to the neuregulin family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB005060; BAA23417.1; -.  
 DR EMBL; AF119162; AAF28848.1; -.  
 DR EMBL; AF119151; AAF28848.1; JOINED.  
 DR EMBL; AF119152; AAF28848.1; JOINED.  
 DR EMBL; AF119153; AAF28848.1; JOINED.  
 DR EMBL; AF119154; AAF28848.1; JOINED.  
 DR EMBL; AF119155; AAF28848.1; JOINED.  
 DR EMBL; AF119158; AAF28848.1; JOINED.  
 DR EMBL; AF119159; AAF28848.1; JOINED.  
 DR EMBL; AF119160; AAF28848.1; JOINED.  
 DR EMBL; AF119161; AAF28848.1; JOINED.  
 DR EMBL; AF119162; AAF28849.1; -.  
 DR EMBL; AF119151; AAF28849.1; JOINED.  
 DR EMBL; AF119152; AAF28849.1; JOINED.  
 DR EMBL; AF119153; AAF28849.1; JOINED.  
 DR EMBL; AF119154; AAF28849.1; JOINED.  
 DR EMBL; AF119156; AAF28849.1; JOINED.  
 DR EMBL; AF119158; AAF28849.1; JOINED.  
 DR EMBL; AF119159; AAF28849.1; JOINED.  
 DR EMBL; AF119160; AAF28849.1; JOINED.  
 DR EMBL; AF119161; AAF28849.1; JOINED.  
 DR EMBL; AF119162; AAF28850.1; -.  
 DR EMBL; AF119151; AAF28850.1; JOINED.  
 DR EMBL; AF119152; AAF28850.1; JOINED.  
 DR EMBL; AF119153; AAF28850.1; JOINED.  
 DR EMBL; AF119154; AAF28850.1; JOINED.  
 DR EMBL; AF119155; AAF28850.1; JOINED.  
 DR EMBL; AF119157; AAF28850.1; JOINED.  
 DR EMBL; AF119158; AAF28850.1; JOINED.  
 DR EMBL; AF119159; AAF28850.1; JOINED.  
 DR EMBL; AF119160; AAF28850.1; JOINED.  
 DR EMBL; AF119161; AAF28850.1; JOINED.  
 DR EMBL; AF119162; AAF28851.1; -.  
 DR EMBL; AF119151; AAF28851.1; JOINED.  
 DR EMBL; AF119152; AAF28851.1; JOINED.  
 DR EMBL; AF119153; AAF28851.1; JOINED.  
 DR EMBL; AF119154; AAF28851.1; JOINED.  
 DR EMBL; AF119156; AAF28851.1; JOINED.  
 DR EMBL; AF119157; AAF28851.1; JOINED.  
 DR EMBL; AF119158; AAF28851.1; JOINED.  
 DR EMBL; AF119159; AAF28851.1; JOINED.  
 DR EMBL; AF119160; AAF28851.1; JOINED.  
 DR EMBL; AF119161; AAF28851.1; JOINED.  
 DR EMBL; AF119158; AAF28852.1; -.

DR EMBL; AF119151; AAF28852.1; JOINED.  
 DR EMBL; AF119152; AAF28852.1; JOINED.  
 DR EMBL; AF119153; AAF28852.1; JOINED.  
 DR EMBL; AF119154; AAF28852.1; JOINED.  
 DR EMBL; AF119155; AAF28852.1; JOINED.  
 DR EMBL; AF119156; AAF28852.1; JOINED.  
 DR EMBL; AF119157; AAF28853.1; -.  
 DR EMBL; AF119151; AAF28853.1; JOINED.  
 DR EMBL; AF119152; AAF28853.1; JOINED.  
 DR EMBL; AF119153; AAF28853.1; JOINED.  
 DR EMBL; AF119154; AAF28853.1; JOINED.  
 DR EMBL; AF119155; AAF28853.1; JOINED.  
 DR EMBL; AF119156; AAF28853.1; JOINED.  
 DR PIR; JC5700; JC5700.  
 DR HSSP; Q12784; 1HRE.  
 DR Genew; HGNC:7998; NRG2.  
 DR MIM; 603818; -.  
 DR GO; GO:0005102; F:receptor binding; TAS.  
 DR GO; GO:0007165; P:signal transduction; TAS.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR InterPro; IPR002154; Neuregulin.  
 DR Pfam; PF00008; EGF; 1.  
 DR Pfam; PF00047; ig; 1.  
 DR Pfam; PF02158; Neuregulin; 1.  
 DR SMART; SM00181; EGF; 1.  
 DR SMART; SM00408; IGc2; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS50026; EGF\_3; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 KW Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;  
 KW Transmembrane; Multigene family; Alternative splicing.  
 FT PROPEP 1 111 BY SIMILARITY.  
 FT CHAIN 112 850 PRO-NEUREGULIN-2, MEMBRANE-BOUND FORM.  
 FT CHAIN 112 404 NEUREGULIN-2.  
 FT DOMAIN 112 405 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 406 426 INTERNAL SIGNAL SEQUENCE (POTENTIAL).  
 FT DOMAIN 427 850 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 237 332 IG-LIKE C2-TYPE.  
 FT DOMAIN 330 340 SER/THR-RICH.  
 FT DOMAIN 341 382 EGF-LIKE.  
 FT DOMAIN 10 13 POLY-PRO.  
 FT DOMAIN 20 30 POLY-SER.  
 FT DOMAIN 33 47 POLY-SER.  
 FT DOMAIN 87 90 POLY-ALA.  
 FT DOMAIN 721 727 POLY-PRO.  
 FT DISULFID 257 311 BY SIMILARITY.  
 FT DISULFID 345 359 BY SIMILARITY.  
 FT DISULFID 353 370 BY SIMILARITY.  
 FT DISULFID 372 381 BY SIMILARITY.  
 FT CARBOHYD 52 52 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 278 278 N-LINKED (GLCNAC. . .) (POTENTIAL).



FT	CARBOHYD	346	346	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	1	233	MRQVCCSALPPPPLEKGRCSSYSDSSSSSSSERSSSSSSSSSS
FT				ESGSSSRSSSNSSISRPAAPPEPRPQQQPQPRSPAARRAA
FT				ARSRAAAAGGMRRDPAPGFSMLLFGVSLACYSPLKSVQDQ
FT				AYKAPVVVEGKVQGLVPAGGSSSNSTREPPASGRVALVKVL
FT				DKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPL
FT				VFKTAFAPLDTNGKNLKKEVGKILCTDC -> MSESRRRGR

Query Match 92.1%; Score 1610; DB 1; Length 850;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-121;  
 Matches 304; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP	60
Db	93	MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP	152
Qy	61	PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFE	120
Db	153	PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFE	212
Qy	121	PLDTNGKNLKKEVGKILCTDCATRPKLKMKMSQTGQVGEKQSLKCEAAAGNPQPSYRWFK	180
Db	213	PLDTNGKNLKKEVGKILCTDCATRPKLKMKMSQTGQVGEKQSLKCEAAAGNPQPSYRWFK	272
Qy	181	DGKELNRSRDIRIKYGNRKNLQFNKVKVEDAGEYVCEAENILGKDTVGRGRLYVNSVS	240
Db	273	DGKELNRSRDIRIKYGNRKNLQFNKVKVEDAGEYVCEAENILGKDTVGRGRLYVNSVS	332
Qy	241	TTLSSWSGHARKCNETAKSYCVNGGVCIYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP	300
Db	333	TTLSSWSGHARKCNETAKSYCVNGGVCIYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP	392
Qy	301	DPKQ	304
Db	393	DPKQ	396

## RESULT 2

### NRG2\_RAT

ID NRG2\_RAT STANDARD; PRT; 868 AA.  
 AC 035569; 035073; 035570; 035571; 035572;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Pro-neuregulin-2 precursor (Pro-NRG2) [Contains: Neuregulin-2 (NRG-2)  
 DE (Neural-and thymus-derived activator for ERBB kinases) (NTAK)].  
 GN NRG2 OR NTAK.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A., SEQUENCE OF 128-162, AND ALTERNATIVE SPLICING.  
 RX MEDLINE=98006324; PubMed=9348101;  
 RA Higashiyama S., Horikawa M., Yamada K., Ichino N., Nakano N.,  
 RA Nakagawa T., Miyagawa J., Matsushita N., Nagatsu T., Taniguchi N.,  
 RA Ishiguro H.;

RT "A novel brain-derived member of the epidermal growth factor family  
RT that interacts with ErbB3 and ErbB4.";  
RL J. Biochem. 122:675-680(1997).  
RN [2]  
RP SEQUENCE OF 109-868 FROM N.A. (ISOFORMS 6 AND 7).  
RC TISSUE=Cerebellum;  
RX MEDLINE=97311397; PubMed=9168114;  
RA Chang H., Riese D.J. II, Gilbert W., Stern D.F., McMahon U.J.;  
RT "Ligands for ErbB-family receptors encoded by a neuregulin-like  
RT gene.";  
RL Nature 387:509-512(1997).  
CC -!- FUNCTION: Direct ligand for ERBB3 and ERBB4 tyrosine kinase  
CC receptors. Concomitantly recruits ERBB1 and ERBB2 coreceptors,  
CC resulting in ligand-stimulated tyrosine phosphorylation and  
CC activation of the ERBB receptors. May also promote the  
CC heterodimerization with the EGF receptor.  
CC -!- SUBCELLULAR LOCATION: EXISTS AS AN TYPE I MEMBRANE PROTEIN AND AS  
CC A PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE  
CC MEMBRANE-BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=7;  
CC Comment=Additional isoforms seem to exist. The alpha-type and  
CC beta-type differ in the EGF-LIKE domain;  
CC Name=1; Synonyms=NTAK-alpha1;  
CC IsoId=035569-1; Sequence=Displayed;  
CC Name=2; Synonyms=NTAK-alpha2A;  
CC IsoId=035569-2; Sequence=VSP\_003471;  
CC Name=3; Synonyms=NTAK-alpha2B, NTAK-alpha2-1P;  
CC IsoId=035569-3; Sequence=VSP\_003466, VSP\_003471;  
CC Name=4; Synonyms=NTAK-beta;  
CC IsoId=035569-4; Sequence=VSP\_003470;  
CC Name=5; Synonyms=NTAK-gamma;  
CC IsoId=035569-5; Sequence=VSP\_003467, VSP\_003468;  
CC Name=6; Synonyms=NRG2-alpha;  
CC IsoId=035569-6; Sequence=VSP\_003472, VSP\_003473;  
CC Name=7; Synonyms=NRG2-beta;  
CC IsoId=035569-7; Sequence=VSP\_003465, VSP\_003469;  
CC -!- TISSUE SPECIFICITY: Expressed in most parts of the brain,  
CC especially the olfactory bulb and cerebellum where it localizes in  
CC granule and Purkinje cells. In the hippocampus, found in the  
CC granule cells of the dentate gyrus. In the basal forebrain, found  
CC in the cholinergic cells. In the hindbrain, weakly detectable in  
CC the motor trigeminal nucleus. Not detected in the hypothalamus.  
CC Also found in the liver and in the thymus. Not detected in heart,  
CC adrenal gland, or testis.  
CC -!- DEVELOPMENTAL STAGE: In the embryo, expressed in the brain of  
CC ell.5 embryos where it is found in the telencephalon, but not in  
CC the hindbrain. Not found in the heart. In the adult, found in  
CC brain and thymus.  
CC -!- DOMAIN: The cytoplasmic domain may be involved in the regulation  
CC of trafficking and proteolytic processing. Regulation of the  
CC proteolytic processing involves initial intracellular domain  
CC dimerization (By similarity).  
CC -!- DOMAIN: ERBB receptor binding is elicited entirely by the EGF-like  
CC domain (By similarity).  
CC -!- PTM: Proteolytic cleavage close to the plasma membrane on the  
CC external face leads to the release of the soluble growth factor

```

CC      form (By similarity).
CC      -!- PTM: Extensive glycosylation precedes the proteolytic cleavage (By
CC      similarity).
CC      -!- SIMILARITY: Contains 1 EGF-like domain.
CC      -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
CC      -!- SIMILARITY: Belongs to the neuregulin family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; D89995; BAA23344.1; -.
DR      EMBL; D89996; BAA23345.1; -.
DR      EMBL; D89997; BAA23346.1; -.
DR      EMBL; D89998; BAA23347.1; -.
DR      EMBL; AB001576; BAA23348.1; -.
DR      PIR; JC5701; JC5701.
DR      PIR; JC5702; JC5702.
DR      HSSP; Q12784; 1HRE.
DR      InterPro; IPR006209; EGF_like.
DR      InterPro; IPR006210; IEGF.
DR      InterPro; IPR007110; Ig-like.
DR      InterPro; IPR003598; Ig_c2.
DR      InterPro; IPR002154; Neuregulin.
DR      Pfam; PF00008; EGF; 1.
DR      Pfam; PF00047; ig; 1.
DR      Pfam; PF02158; Neuregulin; 1.
DR      SMART; SM00181; EGF; 1.
DR      SMART; SM00408; IGc2; 1.
DR      PROSITE; PS00022; EGF_1; 1.
DR      PROSITE; PS01186; EGF_2; 1.
DR      PROSITE; PS50026; EGF_3; 1.
DR      PROSITE; PS50835; IG_LIKE; 1.
KW      Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;
KW      Transmembrane; Multigene family; Alternative splicing.
FT      PROPEP      1      127
FT      CHAIN      128      868      PRO-NEUREGULIN-2, MEMBRANE-BOUND FORM.
FT      CHAIN      128      428      NEUREGULIN-2.
FT      DOMAIN     128      429      EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM   430      450      INTERNAL SIGNAL SEQUENCE (POTENTIAL).
FT      DOMAIN     451      868      CYTOPLASMIC (POTENTIAL).
FT      DOMAIN     253      348      IG-LIKE C2-TYPE.
FT      DOMAIN     346      356      SER/THR-RICH.
FT      DOMAIN     357      398      EGF-LIKE.
FT      DOMAIN     22       32      POLY-SER.
FT      DOMAIN     35       45      POLY-SER.
FT      DOMAIN     56       59      POLY-THR.
FT      DOMAIN     103      106      POLY-ALA.
FT      DOMAIN     739      745      POLY-PRO.
FT      DISULFID   273      327      BY SIMILARITY.
FT      DISULFID   361      375      BY SIMILARITY.
FT      DISULFID   369      386      BY SIMILARITY.
FT      DISULFID   388      397      BY SIMILARITY.

```

FT	CARBOHYD	33	33	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	34	34	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	163	163	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	294	294	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	362	362	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	1	108	Missing (in isoform 7).
FT				/FTId=VSP_003465.
FT	VARSPLIC	220	222	PLV -> FFF (in isoform 3).
FT				/FTId=VSP_003466.
FT	VARSPLIC	388	388	C -> G (in isoform 5).
FT				/FTId=VSP_003467.
FT	VARSPLIC	389	868	Missing (in isoform 5).
FT				/FTId=VSP_003468.
FT	VARSPLIC	390	412	NGFFGQRCLEKLPLRLYMPDPKQ -> VGYTGDRCCQQFAMV
FT				NFS (in isoform 7).
FT				/FTId=VSP_003469.
FT	VARSPLIC	390	421	NGFFGQRCLEKLPLRLYMPDPKQKHLGFELKE -> VGYTG
FT				DRCQQFAMVNFSK (in isoform 4).
FT				/FTId=VSP_003470.
FT	VARSPLIC	414	421	Missing (in isoform 2 and isoform 3).
FT				/FTId=VSP_003471.
FT	VARSPLIC	414	439	HLCGFELKEAEELYQKRVLTITGICVA -> SVLWDTPGTGV
FT				SSSQWSTSPSTLDLN (in isoform 6).
FT				/FTId=VSP_003472.
FT	VARSPLIC	440	868	Missing (in isoform 6).
FT				/FTId=VSP_003473.
FT	CONFLICT	117	117	S -> F (IN REF. 2).
FT	CONFLICT	724	724	R -> H (IN REF. 2).
SQ	SEQUENCE	868 AA;	93776 MW;	3C7D4D94DBE64DE2 CRC64;

Query Match 90.1%; Score 1576; DB 1; Length 868;

Best Local Similarity 96.7%; Pred. No. 9.5e-119;

Matches 297; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY	1	MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP	60
Db	109	MRRDPAPGSSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLAPAGGSSSNSTREP	168
QY	61	PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA	120
Db	169	PASGRVALVKVLDKWPLRSGGLQREQVISVGSCAPLERNQRYIFFLEPTEQPLVFKTAFA	228
QY	121	PLDTNGKNLKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK	180
		:      :	
Db	229	PVDPNGKNIKKEVGKILCTDCATRPKLKKMKSQTGGEVGEKQSLKCEAAAGNPQPSYRWFK	288
QY	181	DGKELNRSRDIRIKYGNRKNLSRLQFNKVKVEDAGEYVCEAENILGKDTVGRRLVNSVS	240
Db	289	DGKELNRSRDIRIKYGNRKNLSRLQFNKVKVEDAGEYVCEAENILGKDTVGRRLHVNSVS	348
QY	241	TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP	300
Db	349	TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP	408
QY	301	DPKQSVL	307
Db	409	DPKQKHL	415

# RESULT 3

## NRG2\_MOUSE

ID NRG2\_MOUSE STANDARD; PRT; 756 AA.  
AC P56974;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Pro-neuregulin-2 precursor (Pro-NRG2) [Contains: Neuregulin-2 (NRG-2)  
DE (Divergent of neuregulin 1) (DON-1)].  
GN NRG2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS NRG2-5; NRG2-10 AND NRG2-16A).  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RX MEDLINE=97311398; PubMed=9168115;  
RA Carraway K.L. III, Weber J.L., Unger M.J., Ledesma J., Yu N.,  
RA Gassmann M., Lai C.;  
RT "Neuregulin-2, a new ligand of ErbB3/ErbB4-receptor tyrosine  
RT kinases.";  
RL Nature 387:512-516(1997).  
RN [2]  
RP SEQUENCE OF 150-756 FROM N.A. (ISOFORMS DON-1M AND DON-1S).  
RC TISSUE=Choroid plexus;  
RX MEDLINE=97342638; PubMed=9199335;  
RA Busfield S.J., Michnick D.A., Chickering T.W., Revett T.L., Ma J.,  
RA Woolf E.A., Comrack C.A., Dussault B.J., Woolf J., Goodearl A.D.J.,  
RA Gearing D.P.;  
RT "Characterization of a neuregulin-related gene, Don-1, that is highly  
RT expressed in restricted regions of the cerebellum and hippocampus.";  
RL Mol. Cell. Biol. 17:4007-4014(1997).  
CC -!- FUNCTION: Direct ligand for ERBB3 and ERBB4 tyrosine kinase  
CC receptors. Concomitantly recruits ERBB1 and ERBB2 coreceptors,  
CC resulting in ligand-stimulated tyrosine phosphorylation and  
CC activation of the ERBB receptors. May also promote the  
CC heterodimerization with the EGF receptor.  
CC -!- SUBCELLULAR LOCATION: EXISTS AS AN TYPE I MEMBRANE PROTEIN AND AS  
CC A PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE  
CC MEMBRANE-BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=4;  
CC Comment=Additional isoforms seem to exist;  
CC Name=NRG2-16A;  
CC IsoId=P56974-1; Sequence=Displayed;  
CC Name=DON-1M;  
CC IsoId=P56974-2; Sequence=VSP\_003464;  
CC Name=DON-1S; Synonyms=NRG2-5;  
CC IsoId=P56974-3; Sequence=VSP\_003462, VSP\_003463;  
CC Name=NRG2-10;  
CC IsoId=P56974-4; Sequence=VSP\_003460, VSP\_003461;  
CC -!- TISSUE SPECIFICITY: Highest expression in the brain, with lower  
CC levels in the lung. In the cerebellum, found in granule and  
CC Purkinje cells.

CC -!- DOMAIN: The cytoplasmic domain may be involved in the regulation  
 CC of trafficking and proteolytic processing. Regulation of the  
 CC proteolytic processing involves initial intracellular domain  
 CC dimerization (By similarity).  
 CC -!- DOMAIN: ERBB receptor binding is elicited entirely by the EGF-like  
 CC domain (By similarity).  
 CC -!- PTM: Proteolytic cleavage close to the plasma membrane on the  
 CC external face leads to the release of the soluble growth factor  
 CC form (By similarity).  
 CC -!- PTM: Extensive glycosylation precedes the proteolytic cleavage (By  
 CC similarity).  
 CC -!- SIMILARITY: Contains 1 EGF-like domain.  
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.  
 CC -!- SIMILARITY: Belongs to the neuregulin family.  
 DR HSSP; Q12784; 1HRE.  
 DR MGD; MGI:1098246; Nrg2.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR InterPro; IPR002154; Neuregulin.  
 DR Pfam; PF00008; EGF; 1.  
 DR Pfam; PF00047; ig; 1.  
 DR Pfam; PF02158; Neuregulin; 1.  
 DR SMART; SM00181; EGF; 1.  
 DR SMART; SM00408; IGc2; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS50026; EGF\_3; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 KW Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;  
 KW Transmembrane; Multigene family; Alternative splicing.  
 FT PROPEP 1 19 BY SIMILARITY.  
 FT CHAIN 20 756 PRO-NEUREGULIN-2, MEMBRANE-BOUND FORM.  
 FT CHAIN 20 314 NEUREGULIN-2.  
 FT DOMAIN 20 315 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 316 336 INTERNAL SIGNAL SEQUENCE (POTENTIAL).  
 FT DOMAIN 337 756 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 145 240 IG-LIKE C2-TYPE.  
 FT DOMAIN 238 248 SER/THR-RICH.  
 FT DOMAIN 249 290 EGF-LIKE.  
 FT DOMAIN 627 633 POLY-PRO.  
 FT DISULFID 165 219 BY SIMILARITY.  
 FT DISULFID 253 267 BY SIMILARITY.  
 FT DISULFID 261 278 BY SIMILARITY.  
 FT DISULFID 280 289 BY SIMILARITY.  
 FT CARBOHYD 55 55 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 254 254 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 280 280 C -> G (in isoform NRG2-10).  
 FT /FTId=VSP\_003460.  
 FT VARSPLIC 281 756 Missing (in isoform NRG2-10).  
 FT /FTId=VSP\_003461.  
 FT VARSPLIC 282 330 VGYTGDRCQQFAMVNFSKHLGFELKEAEELYQKRVLTITGI  
 FT CVALLVVG -> NGFFGQRCLEKLPLRLYMPDPKQSVLWDT  
 FT PGTGVSSSQWSTSPSTLDLN (in isoform DON-1S).

FT				/FTId=VSP_003462.
FT	VARSPLIC	331	756	Missing (in isoform DON-1S).
FT				/FTId=VSP_003463.
FT	VARSPLIC	282	307	VGYTGDRCCQFAMVNFSKHLGFELKE -> NGFFGQRCLEK
FT				LPLRLYMPDPKQK (in isoform DON-1M).
FT				/FTId=VSP_003464.
SQ	SEQUENCE	756 AA;	82213 MW;	51D85DC918BE678E CRC64;

Qy	1	MRRDPAPGFSMLLFVSVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP	60
Db	1	MRRDPAPGFSMLLFVSVSLACYSPLKSVQDQAYKAPVVVEGKVQGLAPAGGSSSNSTREP	60
Qy	61	PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA	120
Db	61	PASGRVALVKVLDKWPLRSGGLQREQVISVGSCAPLERNQRYIFFLEPTEQPLVFKTAFA	120
Qy	121	PLDTNGKNLKEVGKILCTDCATRPKLKKMKSQTGQVGEEKQSLKCEAAAGNPQPSYRWFK	180
		:     :	
Db	121	PVDPNGKNIKKEVGKILCTDCATRPKLKKMKSQTGEVGEKQSLKCEAAAGNPQPSYRWFK	180
Qy	181	DGKELNRSRDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRRLYVNSVS	240
Db	181	DGKELNRSRDIRIKYGNRKN SRLQFNKVRVEDAGEYVCEAENILGKDTVGRRLHVNSVS	240
Qy	241	TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRC	289
Db	241	TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDR	289

## RESULT 4

```

ID      NRGL1_XENLA          STANDARD;          PRT;          677 AA.
AC      O93383; Q9W6N0;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      15-MAR-2004 (Rel. 43, Last annotation update)
DE      Pro-neuregulin-1 precursor (Pro-NRGL) [Contains: Neuregulin-1].
GN      NRGL1.
OS      Xenopus laevis (African clawed frog).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC      Xenopodinae; Xenopus.
OX      NCBI_TaxID=8355;
RN      [1]
RP      SEQUENCE FROM N.A. (ISOFORM ALPHA1), AND ALTERNATIVE SPLICING.
RX      MEDLINE=98352126; PubMed=9685585;
RA      Yang J.F., Zhou H., Pun S., Ip N.Y., Peng H.B., Tsim K.W.K.;
RT      "Cloning of cDNAs encoding xenopus neuregulin: expression in myotomal
RT      muscle during embryo development.";
RL      Brain Res. Mol. Brain Res. 58:59-73(1998).
RN      [2]
RP      SEQUENCE FROM N.A. (ISOFORM CRD).
RX      MEDLINE=99316087; PubMed=10383827;

```

RA Yang J.F., Zhou H., Choi R.C., Ip N.Y., Peng H.B., Tsim K.W.K.;  
 RT "A cysteine-rich form of Xenopus neuregulin induces the expression of  
 RT acetylcholine receptors in cultured myotubes."  
 RL Mol. Cell. Neurosci. 13:415-429(1999).  
 CC -!- FUNCTION: Direct ligand for the ERBB tyrosine kinase receptors.  
 CC Induces expression of acetylcholine receptor in synaptic nuclei.  
 CC -!- SUBCELLULAR LOCATION: Exists as a type I membrane protein and as a  
 CC proteolytically released soluble growth factor form. The membrane-  
 CC bound form does not seem to be active (By similarity).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Comment=Additional isoforms seem to exist. Isoforms have alpha  
 CC or beta-type EGF-like domains;  
 CC Name=Alpha1;  
 CC IsoId=O93383-1; Sequence=Displayed;  
 CC Name=CRD; Synonyms=CRD-NRG1, Cysteine rich domain;  
 CC IsoId=O93383-2; Sequence=VSP\_003449, VSP\_003450;  
 CC -!- TISSUE SPECIFICITY: Isoform alpha1 is expressed in brain and  
 CC muscle. Isoform CRD is expressed in brain and spinal cord, but at  
 CC very low level in muscle.  
 CC -!- DEVELOPMENTAL STAGE: Strong expression in developing brain and  
 CC spinal cord of the embryo. Also expressed in the myotomal muscle.  
 CC -!- DOMAIN: The cytoplasmic domain may be involved in the regulation  
 CC of trafficking and proteolytic processing. Regulation of the  
 CC proteolytic processing involves initial intracellular domain  
 CC dimerization (By similarity).  
 CC -!- DOMAIN: ERBB receptor binding is elicited entirely by the EGF-like  
 CC domain.  
 CC -!- PTM: Proteolytic cleavage close to the plasma membrane on the  
 CC external face leads to the release of the soluble growth factor  
 CC form.  
 CC -!- PTM: Extensive glycosylation precedes the proteolytic cleavage (By  
 CC similarity).  
 CC -!- SIMILARITY: Contains 1 EGF-like domain.  
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.  
 CC -!- SIMILARITY: Belongs to the neuregulin family.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AF076618; AAC26804.1; -.  
 DR EMBL; AF142632; AAD33893.1; -.  
 DR HSSP; Q12784; 1HRE.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR InterPro; IPR002154; Neuregulin.  
 DR Pfam; PF00008; EGF; 1.  
 DR Pfam; PF00047; ig; 1.  
 DR Pfam; PF02158; Neuregulin; 1.  
 DR PRINTS; PR01089; NEUREGULIN.



DR SMART; SM00181; EGF; 1.  
 DR SMART; SM00408; IGc2; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS50026; EGF\_3; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 KW Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;  
 KW Transmembrane; Alternative splicing.  
 FT CHAIN 1 259 NEUREGULIN ALPHA1 (BY SIMILARITY).  
 FT CHAIN 1 677 PRO-NEUREGULIN ALPHA1, MEMBRANE-BOUND  
 FT FORM (BY SIMILARITY).  
 FT DOMAIN 1 260 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 261 280 INTERNAL SIGNAL SEQUENCE (POTENTIAL).  
 FT DOMAIN 281 677 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 37 132 IG-LIKE C2-TYPE.  
 FT DOMAIN 188 232 EGF-LIKE.  
 FT DISULFID 57 116 BY SIMILARITY.  
 FT DISULFID 192 206 BY SIMILARITY.  
 FT DISULFID 200 220 BY SIMILARITY.  
 FT DISULFID 222 231 BY SIMILARITY.  
 FT DOMAIN 1 25 LYS-RICH.  
 FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 1 136 MAEKKKVKEGKGRKGKGGKDRKGKKAEGSDQGAAASPCLKE  
 FT IKTSQSVQEGKKLVLCQAVSEQPSLKFRWFKGEKEIGAKNK  
 FT PDSKPEHIKIRGKKKSSSELQISKASSADNGEYKCMVSNQLG  
 FT NDTVTNVNTIVPK -> MSEDTAEGQLNQCSSEQSSDPPSAE  
 FT LQNEESMPETQDEEETHGITGLAITCCVCLEADRLRICLN  
 FT SEKICIIPILACLISLCLCIAGLKWVFVDKIFEYDSPTHLD  
 FT PGHRGQDLILYTDTPSTLVPSSVRTLPVVIPTTDSKAAVT  
 FT FKFGTSLLPTE (in isoform CRD).  
 FT /FTId=VSP\_003449.  
 FT VARSPLIC 223 252 KPGFTGARCTETDPLRVVRSEKHLGIEFME -> PNEFTGD  
 FT RCQNYVMASFYK (in isoform CRD).  
 FT /FTId=VSP\_003450.  
 SQ SEQUENCE 677 AA; 75794 MW; 49279E8F5BAE396F CRC64;

Query Match 17.8%; Score 311.5; DB 1; Length 677;  
 Best Local Similarity 34.7%; Pred. No. 2.1e-17;  
 Matches 78; Conservative 23; Mismatches 71; Indels 53; Gaps 5;

QY 126 GKNLKKEVGKIL---CTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDG 182  
 || | || | |||::|::| |::| | : | :|||  
 Db 15 GKGKKDRKGKKAEGSDQGAAASPCLKEIKTSQSVQEGKKLVLCQAVSEQPSLKFRWFKGE 74  
 QY 183 KELNR-----SRDIRIKYGNRKNRLQFNKVKVEDAGEYVCEAENILGKDTV----- 230  
 ||: | : || :|:| || :| | || | | || |||  
 Db 75 KEIGAKNKPDSKPEHIKIRGKKKSSSELQISKASSADNGEYKCMVSNQLGNDTVTVNVNTIV 134  
 QY 231 -----RGRLYVNSVSTTL-----SSWSGHARKCNE 255  
 : | | || :: : || ||::  
 Db 135 PKPTYNHLLLMKIYLVKTSVEKSVEPSTLNLLESQKEVIFATTKRGDTTAGPGHLLIKCSD 194  
 QY 256 TAKSYCVNGGVCYYIEGI---NQLSCKCPNGFFGQRCLEKLPLRL 297  
 |:||||| || : || || || | || | |||:  
 Db 195 KEKTYCVNGGECYVLNGITSSNQFMCKCKPGFTGARCTETDPLRV 239

# RESULT 5

## NRG1\_HUMAN

ID NRG1\_HUMAN STANDARD; PRT; 639 AA.  
AC Q02297; Q14667; P98202; Q02298; Q02299; Q07110; Q07111; Q12779;  
AC Q12780; Q12781; Q12782; Q12783; Q12784; Q9UPE3;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Pro-neuregulin-1 precursor (Pro-NRG1) [Contains: Neuregulin-1 (Neu  
DE differentiation factor) (Heregulin) (HRG) (Breast cancer cell  
DE differentiation factor p45) (Acetylcholine receptor inducing activity)  
DE (ARIA) (Sensory and motor neuron-derived factor) (Glial growth  
DE factor)].  
GN NRG1 OR HGL OR NDF OR HRGA OR GGF OR SMDF.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1; 6; 7 AND 8), AND PARTIAL  
RP SEQUENCE.  
RX MEDLINE=92271253; PubMed=1350381;  
RA Holmes W.E., Sliwkowski M.X., Akita R.W., Henzel W.J., Lee J.,  
RA Park J.W., Yansura D., Abadi N., Raab H., Lewis G.D., Shepard H.M.,  
RA Kuang W.-J., Wood W.I., Goeddel D.V., Vandlen R.L.;  
RT "Identification of heregulin, a specific activator of p185erbB2.";  
RL Science 256:1205-1210(1992).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4; 6; 7 AND 8).  
RC TISSUE=Kidney adenocarcinoma, and Pituitary;  
RX MEDLINE=94158863; PubMed=7509448;  
RA Wen D., Suggs S.V., Karunakaran D., Liu N., Cupples R.L., Luo Y.,  
RA Janssen A.M., Ben-Baruch N., Trollinger D.B., Jacobsen V.L.,  
RA Meng S.-Y., Lu H.S., Hu S., Chang D., Yang W., Yanigahara D.,  
RA Koski R.A., Yarden Y.;  
RT "Structural and functional aspects of the multiplicity of Neu  
RT differentiation factors.";  
RL Mol. Cell. Biol. 14:1909-1919(1994).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=92208945; PubMed=1348215;  
RA Peles E., Bacus S.S., Koski R.A., Lu H.S., Wen D., Ogden S.G.,  
RA Levy R.B., Yarden Y.;  
RT "Isolation of the neu/HER-2 stimulatory ligand: a 44 kd glycoprotein  
RT that induces differentiation of mammary tumor cells.";  
RL Cell 69:205-216(1992).  
RN [4]  
RP SEQUENCE FROM N.A. (ISOFORMS 8 AND 9).  
RC TISSUE=Brain;  
RX MEDLINE=93205115; PubMed=8096067;  
RA Marchionni M.A., Goodearl A.D.J., Chen M.S., Bermingham-McDonogh O.,  
RA Kirk C., Hendricks M., Danehy F., Misumi D., Sudhalter J.,  
RA Kobayashi K., Wroblewski D., Lynch C., Baldassarre M., Hiles I.,  
RA Davis J.B., Hsuan J.J., Totty N.F., Otsu M., McBurney R.N.,  
RA Waterfield M.D., Stroobant P., Gwynne D.;  
RT "Glial growth factors are alternatively spliced erbB2 ligands

RT expressed in the nervous system.";  
 RL Nature 362:312-318(1993).  
 RN [5]  
 RP SEQUENCE FROM N.A. OF GAMMA-HEREGULIN FUSION PROTEIN.  
 RC TISSUE=Breast cancer;  
 RX MEDLINE=97472144; PubMed=9333014;  
 RA Schaefer G., Fitzpatrick V.D., Sliwkowski M.X.;  
 RT "Gamma-heregulin: a novel heregulin isoform that is an autocrine  
 RT growth factor for the human breast cancer cell line, MDA-MB-175.";  
 RL Oncogene 15:1385-1394(1997).  
 RN [6]  
 RP SEQUENCE OF 1-210 FROM N.A.  
 RA Schoumacher F., Herzer S., Flury N., Kueng W., Mueller H.,  
 RA Eppenberger U.;  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.  
 RN [7]  
 RP SEQUENCE OF 19-27.  
 RX MEDLINE=93366731; PubMed=7689552;  
 RA Culouscou J.-M., Plowman G.D., Carlton G.W., Green J.M., Shoyab M.;  
 RT "Characterization of a breast cancer cell differentiation factor that  
 RT specifically activates the HER4/p180erbB4 receptor.";  
 RL J. Biol. Chem. 268:18407-18410(1993).  
 RN [8]  
 RP CHROMOSOMAL TRANSLOCATION.  
 RX MEDLINE=99455251; PubMed=10523851;  
 RA Wang X.-Z., Jolicoeur E.M., Conte N., Chaffanet M., Zhang Y.,  
 RA Mozziconacci M.-J., Feiner H., Birnbaum D., Pebusque M.-J., Ron D.;  
 RT "Gamma-heregulin is the product of a chromosomal translocation fusing  
 RT the DOC4 and HGL/NRG1 genes in the MDA-MB-175 breast cancer cell  
 RT line.";  
 RL Oncogene 18:5718-5721(1999).  
 RN [9]  
 RP CHROMOSOMAL TRANSLOCATION.  
 RX MEDLINE=20065180; PubMed=10597312;  
 RA Liu X., Baker E., Eyre H.J., Sutherland G.R., Zhou M.;  
 RT "Gamma-heregulin: a fusion gene of DOC-4 and neuregulin-1 derived from  
 RT a chromosome translocation.";  
 RL Oncogene 18:7110-7114(1999).  
 RN [10]  
 RP STRUCTURE BY NMR OF 175-241 (ISOFORM 1).  
 RX MEDLINE=94341264; PubMed=8062828;  
 RA Nagata K., Kohda D., Hatanaka H., Ichikawa S., Matsuda S.,  
 RA Yamamoto T., Suzuki A., Inagaki F.;  
 RT "Solution structure of the epidermal growth factor-like domain of  
 RT heregulin-alpha, a ligand for p180erbB-4.";  
 RL EMBO J. 13:3517-3523(1994).  
 CC -!- FUNCTION: Direct ligand for ERBB3 and ERBB4 tyrosine kinase  
 CC receptors. Concomitantly recruits ERBB1 and ERBB2 coreceptors,  
 CC resulting in ligand-stimulated tyrosine phosphorylation and  
 CC activation of the ERBB receptors. The multiple isoforms perform  
 CC diverse functions such as inducing growth and differentiation of  
 CC epithelial, glial, neuronal, and skeletal muscle cells; inducing  
 CC expression of acetylcholine receptor in synaptic vesicles during  
 CC the formation of the neuromuscular junction; stimulating  
 CC lobuloalveolar budding and milk production in the mammary gland  
 CC and inducing differentiation of mammary tumor cells; stimulating  
 CC Schwann cell proliferation; implication in the development of the

myocardium such as trabeculation of the developing heart.

-!- SUBUNIT: The cytoplasmic domain interacts with the LIM domain region of LIMK1 (By similarity).

-!- SUBCELLULAR LOCATION: Exists as an type I membrane protein and as a proteolytically released soluble growth factor form. The membrane-bound form does not seem to be active. The secreted isoform 9 has a signal peptide. The isoform 8 may be nuclear.

-!- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=9;

Comment=Additional isoforms seem to exist. Isoforms have been classified as type I NRGS (isoforms with an Ig domain and a glycosylation domain, isoforms 1-8), type II NRGS (isoforms with an Ig domain but no glycosylation domain, isoform 9) and type III NRGS (isoforms with a Cys-rich domain, isoform 10). All these isoforms perform distinct tissue-specific functions;

Name=1; Synonyms=Alpha;

IsoId=Q02297-1; Sequence=Displayed;

Name=2; Synonyms=Alpha1A;

IsoId=Q02297-2; Sequence=VSP\_003431;

Name=3; Synonyms=Alpha2B;

IsoId=Q02297-3; Sequence=VSP\_003434, VSP\_003435;

Name=4; Synonyms=Alpha3;

IsoId=Q02297-4; Sequence=VSP\_003432, VSP\_003433;

Name=6; Synonyms=Beta1, Beta1A;

IsoId=Q02297-6; Sequence=VSP\_003428;

Name=7; Synonyms=Beta2;

IsoId=Q02297-7; Sequence=VSP\_003427;

Name=8; Synonyms=Beta3, GGFHFB1;

IsoId=Q02297-8; Sequence=VSP\_003429, VSP\_003430;

Name=9; Synonyms=GGF2, GGFHPP2;

IsoId=Q02297-9; Sequence=VSP\_003425, VSP\_003426, VSP\_003429, VSP\_003430;

Name=10; Synonyms=SMDF;

IsoId=Q15491-1; Sequence=External;

-!- TISSUE SPECIFICITY: Type I isoforms are the predominant forms expressed in the endocardium. Isoform alpha is expressed in breast, ovary, testis, prostate, heart, skeletal muscle, lung, placenta liver, kidney, salivary gland, small intestine and brain, but not in uterus, stomach, pancreas, and spleen. Isoform 3 is the predominant form in mesenchymal cells and in nonneuronal organs, whereas isoform 5 is the major neuronal form. Isoform 8 is expressed in spinal cord and brain. Isoform 9 is the major form in skeletal muscle cells; in the nervous system it is expressed in spinal cord and brain. Also detected in adult heart, placenta, lung, liver, kidney, and pancreas.

-!- DEVELOPMENTAL STAGE: Detectable at early embryonic ages.

-!- DOMAIN: The cytoplasmic domain may be involved in the regulation of trafficking and proteolytic processing. Regulation of the proteolytic processing involves initial intracellular domain dimerization (By similarity).

-!- DOMAIN: ERBB receptor binding is elicited entirely by the EGF-like domain.

-!- PTM: Proteolytic cleavage close to the plasma membrane on the external face leads to the release of the soluble growth factor form.

-!- PTM: Extensive glycosylation precedes the proteolytic cleavage (By similarity).

CC -!- DISEASE: Involved in a rare t(8;11) chromosomal translocation that  
 CC fuses the 5'end of ODZ4 to NRG1 (isoform 8). The product of this  
 CC translocation was first thought to be an alternatively spliced  
 CC isoform, called gamma-heregulin. Gamma-heregulin is a soluble  
 CC activating ligand for the ERBB2-ERBB3 receptor complex and acts as  
 CC an autocrine growth factor in a specific breast cancer cell line  
 CC (MDA-MB-175). Not detected in breast carcinoma samples, including  
 CC ductal, lobular, medullary, and mucinous histological types,  
 CC neither in other breast cancer cell lines.  
 CC -!- SIMILARITY: Contains 1 EGF-like domain.  
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.  
 CC -!- SIMILARITY: Belongs to the neuregulin family.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; M94165; AAA58638.1; -.  
 DR EMBL; M94166; AAA58639.1; -.  
 DR EMBL; M94167; AAA58640.1; -.  
 DR EMBL; M94168; AAA58641.1; -.  
 DR EMBL; U02325; AAA19950.1; -.  
 DR EMBL; U02326; AAA19951.1; -.  
 DR EMBL; U02327; AAA19952.1; -.  
 DR EMBL; U02328; AAA19953.1; -.  
 DR EMBL; U02329; AAA19954.1; -.  
 DR EMBL; U02330; AAA19955.1; -.  
 DR EMBL; L12260; AAB59622.1; -.

Query Match 16.8%; Score 293.5; DB 1; Length 639;  
 Best Local Similarity 32.1%; Pred. No. 5.6e-16;  
 Matches 71; Conservative 37; Mismatches 62; Indels 51; Gaps 7;

QY 126 GKNLKKEVGKILCTDCAT-----RPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRW 178  
 || ||| | : | |::||| | | |:: : :  
 Db 10 GKGKKKERGSGKKPESAAGSQSPALPPRLKEMKSQESAAGSKLVLCETSSEYSSLRFKW 69  
 QY 179 FKDGKELNRS---RDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGL- 234  
 ||:| ||| :::|: | : | | : |::|: : || | :  
 Db 70 FKNGNELNRKNKPQNIKIQQKPGK--SELRINKASLADSGEYMCKVISKLGNDASANIT 127  
 QY 235 -----YV-----NSVSTTLSSWSG--HARKCNETAKS 259  
 || | : | : | : | | | | :  
 Db 128 IVESNEIITGMPASTEGAYVSSSPIRISVSTEGANTSSSTSTSTGTGSHLVKCAEKEKT 187  
 QY 260 YCVNGGVCYYIEGINQLS---CKCPNGFFGQRCLEKLPRL 297  
 :||| | : : : | | | | | | ::::  
 Db 188 FCVNGGECFMVKDLNPSRYLCKCQPGFTGARCTENVPMKV 228

RESULT 6

NRG1\_CHICK

ID NRG1\_CHICK STANDARD; PRT; 602 AA.

AC Q05199; O73750; O73751; O73752;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Pro-neuregulin-1 precursor (Pro-NRG1) [Contains: Neuregulin-1  
 DE (Acetylcholine receptor inducing activity) (ARIA)].  
 GN NRG1 OR ARIA.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND PARTIAL SEQUENCE.  
 RC STRAIN=White leghorn; TISSUE=Brain;  
 RX MEDLINE=93201602; PubMed=8453670;  
 RA Falls D.L., Rosen K.M., Corfas G., Lane W.S., Fischbach G.D.;  
 RT "ARIA, a protein that stimulates acetylcholine receptor synthesis, is  
 RT a member of the neu ligand family.";  
 RL Cell 72:801-815(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 2; 3 AND 4).  
 RC TISSUE=Brain, and Spinal cord;  
 RX MEDLINE=98150951; PubMed=9491987;  
 RA Yang X., Kuo Y., Devay P., Yu C., Role L.;  
 RT "A cysteine-rich isoform of neuregulin controls the level of  
 RT expression of neuronal nicotinic receptor channels during  
 RT synaptogenesis.";  
 RL Neuron 20:255-270(1998).  
 CC -!- FUNCTION: Direct ligand for the ERBB tyrosine kinase receptors.  
 CC The multiple isoforms perform diverse functions: Cystein-rich  
 CC domain containing isoforms (isoforms 2-4) probably regulate the  
 CC expression of nicotinic acetylcholine receptors at developing  
 CC interneuronal synapses. The Ig-NRG isoform is required for the  
 CC initial induction and/or maintenance of the mature levels of  
 CC acetylcholine receptors at neuromuscular synapses.  
 CC -!- SUBCELLULAR LOCATION: Exists as a type I membrane protein and as a  
 CC proteolytically released soluble growth factor form. The membrane-  
 CC bound form does not seem to be active (By similarity).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=4;  
 CC Comment=Additional isoforms seem to exist;  
 CC Name=1; Synonyms=ARIA, IG-NRG;  
 CC IsoId=Q05199-1; Sequence=Displayed;  
 CC Note=Contains an Ig-like domain;  
 CC Name=2; Synonyms=CRD-NRG-BETA1A;  
 CC IsoId=Q05199-2; Sequence=VSP\_003445;  
 CC Note=The EGF-like domain is replaced by a Cysteine-rich domain  
 CC (CRD);  
 CC Name=3; Synonyms=CRD-NRG-BETA2A;  
 CC IsoId=Q05199-3; Sequence=VSP\_003445, VSP\_003446;  
 CC Note=The EGF-like domain is replaced by a Cysteine-rich domain  
 CC (CRD);  
 CC Name=4; Synonyms=CRD-NRG-BETA2B;  
 CC IsoId=Q05199-4; Sequence=VSP\_003445, VSP\_003446, VSP\_003447,  
 CC VSP\_003448;  
 CC Note=The EGF-like domain is replaced by a Cysteine-rich domain

```

CC      (CRD);
CC      -!- DEVELOPMENTAL STAGE: Isoforms 2-4 are detected at embryonic day 4
CC          (ED4) in both visceral and somatic motor neurons of spinal cord
CC          and is highest at ED6. Isoform 1 is not expressed until ED 6 in
CC          spinal cord. At ED 11 both isoforms display comparable levels.
CC      -!- DOMAIN: The cytoplasmic domain may be involved in the regulation
CC          of trafficking and proteolytic processing. Regulation of the
CC          proteolytic processing involves initial intracellular domain
CC          dimerization (By similarity).
CC      -!- DOMAIN: ERBB receptor binding is elicited entirely by the EGF-like
CC          domain.
CC      -!- PTM: Proteolytic cleavage close to the plasma membrane on the
CC          external face leads to the release of the soluble growth factor
CC          form.
CC      -!- PTM: Extensive glycosylation precedes the proteolytic cleavage (By
CC          similarity).
CC      -!- SIMILARITY: Contains 1 EGF-like domain.
CC      -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
CC      -!- SIMILARITY: Belongs to the neuregulin family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; L11264; AAA49037.1; -.
DR      EMBL; AF045654; AAC05670.1; -.
DR      EMBL; AF045655; AAC05671.1; -.
DR      EMBL; AF045656; AAC05672.1; -.
DR      PIR; A45769; A45769.
DR      HSSP; Q12784; 1HRE.
DR      InterPro; IPR006209; EGF_like.
DR      InterPro; IPR006210; IEGF.
DR      InterPro; IPR007110; Ig-like.
DR      InterPro; IPR003598; Ig_c2.
DR      InterPro; IPR002154; Neuregulin.
DR      Pfam; PF00047; ig; 1.
DR      Pfam; PF02158; Neuregulin; 1.
DR      PRINTS; PR01089; NEUREGULIN.
DR      SMART; SM00181; EGF; 1.
DR      SMART; SM00408; IGc2; 1.
DR      PROSITE; PS00022; EGF_1; 1.
DR      PROSITE; PS01186; EGF_2; FALSE_NEG.
DR      PROSITE; PS50026; EGF_3; 1.
DR      PROSITE; PS50835; IG_LIKE; 1.
KW      Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;
KW      Transmembrane; Alternative splicing.
FT      CHAIN      1      602      PRO-NEUREGULIN-1, MEMBRANE-BOUND FORM.
FT      CHAIN      1      205      NEUREGULIN-1.
FT      DOMAIN     1      206      EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM   207      229      INTERNAL SIGNAL SEQUENCE (POTENTIAL).
FT      DOMAIN     230      602      CYTOPLASMIC (POTENTIAL).
FT      DOMAIN     29      123      IG-LIKE C2-TYPE.
FT      DOMAIN     125      136      SER/THR-RICH.

```





DE differentiation factor) (Heregulin) (HRG) (Acetylcholine receptor  
DE inducing activity) (ARIA) (Sensory and motor neuron-derived factor)  
DE (Glial growth factor)].

GN NRG1 OR NDF.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.

RC TISSUE=Fibroblast;

RX MEDLINE=94158863; PubMed=7509448;

RA Wen D., Suggs S.V., Karunakaran D., Liu N., Cupples R.L., Luo Y.,  
RA Janssen A.M., Ben-Baruch N., Trollinger D.B., Jacobsen V.L.,  
RA Meng S.-Y., Lu H.S., Hu S., Chang D., Yang W., Yanigahara D.,  
RA Koski R.A., Yarden Y.;

RT "Structural and functional aspects of the multiplicity of Neu  
RT differentiation factors.";

RL Mol. Cell. Biol. 14:1909-1919(1994).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORM ALPHA2C/NDF44), AND PARTIAL SEQUENCE.

RC TISSUE=Fibroblast;

RX MEDLINE=92257596; PubMed=1349853;

RA Wen D., Peles E., Cupples R., Suggs S.V., Bacus S.S., Luo Y.,  
RA Trail G., Hu S., Silbiger S.M., Levy R.B., Koski R.A., Lu H.S.,  
RA Yarden Y.;

RT "Neu differentiation factor: a transmembrane glycoprotein containing  
RT an EGF domain and an immunoglobulin homology unit.";

RL Cell 69:559-572(1992).

RN [3]

RP SEQUENCE OF 14-36.

RX MEDLINE=92208945; PubMed=1348215;

RA Peles E., Bacus S.S., Koski R.A., Lu H.S., Wen D., Ogden S.G.,  
RA Levy R.B., Yarden Y.;

RT "Isolation of the neu/HER-2 stimulatory ligand: a 44 kd glycoprotein  
RT that induces differentiation of mammary tumor cells.";

RL Cell 69:205-216(1992).

RN [4]

RP REGULATION OF PROCESSING (ISOFORM ALPHA2C/NDF44).

RX MEDLINE=99069430; PubMed=9852099;

RA Liu X., Hwang H., Cao L., Wen D., Liu N., Graham R.M., Zhou M.;

RT "Release of the neuregulin functional polypeptide requires its  
RT cytoplasmic tail.";

RL J. Biol. Chem. 273:34335-34340(1998).

RN [5]

RP INTERACTION WITH LIMK1.

RX MEDLINE=98352096; PubMed=9685409;

RA Wang J.Y., Frenzel K.E., Wen D., Falls D.L.;

RT "Transmembrane neuregulins interact with LIM kinase 1, a cytoplasmic  
RT protein kinase implicated in development of visuospatial cognition.";

RL J. Biol. Chem. 273:20525-20534(1998).

CC -!- FUNCTION: Direct ligand for ERBB3 and ERBB4 tyrosine kinase  
CC receptors. Concomitantly recruits ERBB1 and ERBB2 coreceptors,  
CC resulting in ligand-stimulated tyrosine phosphorylation and  
CC activation of the ERBB receptors. The multiple isoforms perform  
CC diverse functions such as inducing growth and differentiation of  
CC epithelial, glial, neuronal, and skeletal muscle cells; inducing

CC expression of acetylcholine receptor in synaptic vessicles during  
 CC the formation of the neuromuscular junction; stimulating  
 CC lobuloalveolar budding and milk production in the mammary gland  
 CC and inducing differentiation of mammary tumor cells; stimulating  
 CC Schwann cell proliferation; implication in the development of the  
 CC myocardium such as trabeculation of the developing heart (By  
 CC similarity).  
 CC -!- SUBUNIT: The cytoplasmic domain interacts with the LIM domain  
 CC region of LIMK1.  
 CC -!- SUBCELLULAR LOCATION: Exists as a type I membrane protein and as a  
 CC proteolytically released soluble growth factor form. The membrane-  
 CC bound form does not seem to be active.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=8;  
 CC Comment=Additional isoforms seem to exist;  
 CC Name=Beta4; Synonyms=NDF42A;  
 CC IsoId=P43322-1; Sequence=Displayed;  
 CC Name=Alpha2A; Synonyms=NDF38;  
 CC IsoId=P43322-2; Sequence=VSP\_003436;  
 CC Name=Alpha2B; Synonyms=NDF19;  
 CC IsoId=P43322-3; Sequence=VSP\_003436, VSP\_003443, VSP\_003444;  
 CC Name=Alpha2C; Synonyms=NDF44;  
 CC IsoId=P43322-4; Sequence=VSP\_003436, VSP\_003442;  
 CC Name=Beta1;  
 CC IsoId=P43322-5; Sequence=VSP\_003437;  
 CC Name=Beta2; Synonyms=NDF40;  
 CC IsoId=P43322-6; Sequence=VSP\_003440, VSP\_003441;  
 CC Name=Beta2A; Synonyms=NDF22;  
 CC IsoId=P43322-7; Sequence=VSP\_003440;  
 CC Name=Beta3; Synonyms=NDF4;  
 CC IsoId=P43322-8; Sequence=VSP\_003438, VSP\_003439;  
 CC -!- TISSUE SPECIFICITY: Widely expressed. Most tissues contain alpha2A  
 CC and alpha2B isoforms. Alpha2 and beta2 are the predominant forms  
 CC in mesenchymal and nonneuronal organs. Beta1 is enriched in brain  
 CC and spinal cord, but not in muscle and heart. Alpha2C is highly  
 CC expressed in spinal cord, moderately in lung, brain, ovary, and  
 CC stomach, in low amounts in the kidney, skin and heart and not  
 CC detected in the liver, spleen, and placenta.  
 CC -!- DOMAIN: The cytoplasmic domain may be involved in the regulation of  
 CC trafficking and proteolytic processing. Regulation of the  
 CC proteolytic processing involves initial intracellular domain  
 CC dimerization.  
 CC -!- DOMAIN: ERBB receptor binding is elicited entirely by the EGF-like  
 CC domain.  
 CC -!- PTM: Proteolytic cleavage close to the plasma membrane on the  
 CC external face leads to the release of the soluble growth factor  
 CC form.  
 CC -!- PTM: Extensive glycosylation precedes the proteolytic cleavage.  
 CC -!- SIMILARITY: Contains 1 EGF-like domain.  
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.  
 CC -!- SIMILARITY: Belongs to the neuregulin family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; U02315; AAA19940.1; -.  
 DR EMBL; U02316; AAA19941.1; -.  
 DR EMBL; U02317; AAA19942.1; -.  
 DR EMBL; U02318; AAA19943.1; -.  
 DR EMBL; U02319; AAA19944.1; -.  
 DR EMBL; U02320; AAA19945.1; -.  
 DR EMBL; U02321; AAA19946.1; -.  
 DR EMBL; U02322; AAA19947.1; -.  
 DR EMBL; U02323; AAA19948.1; -.  
 DR EMBL; U02324; AAA19949.1; -.  
 DR EMBL; M92430; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; I61718; I61718.  
 DR PIR; I61719; I61719.  
 DR PIR; I61722; I61722.  
 DR HSSP; Q12784; 1HRE.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR InterPro; IPR002154; Neuregulin.  
 DR Pfam; PF00008; EGF; 1.  
 DR Pfam; PF00047; ig; 1.  
 DR Pfam; PF02158; Neuregulin; 1.  
 DR PRINTS; PR01089; NEUREGULIN.  
 DR SMART; SM00181; EGF; 1.  
 DR SMART; SM00408; IGc2; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; FALSE\_NEG.  
 DR PROSITE; PS50026; EGF\_3; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 KW Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;  
 KW Transmembrane; Multigene family; Alternative splicing.  
 FT PROPEP 1 13  
 FT CHAIN 14 662 PRO-NEUREGULIN-1, MEMBRANE-BOUND FORM.  
 FT CHAIN 14 264 NEUREGULIN-1.  
 FT DOMAIN 14 265 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 266 288 INTERNAL SIGNAL SEQUENCE (POTENTIAL).  
 FT DOMAIN 289 662 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 37 128 IG-LIKE C2-TYPE.  
 FT DOMAIN 165 177 SER/THR-RICH.  
 FT DOMAIN 178 222 EGF-LIKE.  
 FT DISULFID 57 112  
 FT DISULFID 182 196 BY SIMILARITY.  
 FT DISULFID 190 210 BY SIMILARITY.  
 FT DISULFID 212 221 BY SIMILARITY.  
 FT CARBOHYD 120 120 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 213 256 PNEFTGDRQNYVMASFYMTSRRKRQETEKPLERKLDHSLV  
 FT KES -> QPGFTGARCTENVPMKVQTQE (in isoform  
 FT Alpha2A, isoform Alpha2B and isoform  
 FT Alpha2C).  
 FT /FTId=VSP\_003436.  
 FT VARSPLIC 231 257 MTSRRKRQETEKPLERKLDHSLVKESK -> KHLGIEFME



OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC TISSUE=Embryo, and Imaginal disks;  
 RX MEDLINE=96421972; PubMed=8824589;  
 RA Schnepf B.C., Grumbling G.B., Donaldson T.D., Simcox A.A.;  
 RT "Vein is a novel component in the Drosophila epidermal growth factor  
 RT receptor pathway with similarity to the neuregulins.";  
 RL Genes Dev. 10:2302-2313(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Flossler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [3]  
 RP REVISIONS.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,

RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
 RT systematic review.";  
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
 CC -!- FUNCTION: Ligand for the EGF receptor. Seems to play a role in  
 CC the global proliferation of wing disc cells and the larval  
 CC patterning. Shows a strong synergistic genetic interaction with  
 CC spi, suggesting a molecular interdependence. Required for the  
 CC development of interveins cells.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q94918-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q94918-2; Sequence=VSP\_001419;  
 CC -!- DEVELOPMENTAL STAGE: Expressed in blastoderm embryos in two  
 CC ventrolateral stripes that are brought to the midline as  
 CC gastrulation proceeds. In the germ-band retraction stage,  
 CC expression is seen in the CNS and epidermis. At late blastoderm,  
 CC expression is localized in the anlagen of the amnioserosa.  
 CC Expression in the head, cypeolabrum, maxillary and labial lobes,  
 CC and around the stomodeum throughout embryo development. In late  
 CC embryos, expression decays in all ectodermal cells and appears in  
 CC the segmental muscles and the gut wall. In the larva, expression  
 CC occurs in the dorsal metathoracic disc, the eye-antennal disc and  
 CC the ventral thoracic disc. Found in the intervein in the pupa.  
 CC -!- SIMILARITY: Contains 1 EGF-like domain.  
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U67935; AAC47293.1; -.  
 DR EMBL; AE003564; AAF50739.2; -.  
 DR FlyBase; FBgn0003984; vn.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0005154; F:epidermal growth factor receptor binding; IMP.  
 DR GO; GO:0007173; P:EGF receptor signaling pathway; IMP.  
 DR GO; GO:0007477; P:notum morphogenesis; IMP.  
 DR GO; GO:0045742; P:positive regulation of EGF receptor signali. . .; NAS.  
 DR GO; GO:0007476; P:wing morphogenesis; IMP.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR Pfam; PF00008; EGF; 1.

DR Pfam; PF00047; ig; 1.  
 DR SMART; SM00181; EGF; 1.  
 DR SMART; SM00408; IGc2; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; FALSE\_NEG.  
 DR PROSITE; PS50026; EGF\_3; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 KW EGF-like domain; Glycoprotein; Immunoglobulin domain; Growth factor;  
 KW Developmental protein; Alternative splicing; Signal.  
 FT SIGNAL 1 40 POTENTIAL.  
 FT CHAIN 41 623 VEIN PROTEIN.  
 FT DOMAIN 457 542 IG-LIKE C2-TYPE.  
 FT DOMAIN 561 599 EGF-LIKE.  
 FT DOMAIN 43 53 POLY-GLN.  
 FT DOMAIN 87 96 POLY-SER.  
 FT DOMAIN 138 314 GLN-RICH.  
 FT DISULFID 478 531 BY SIMILARITY.  
 FT CARBOHYD 76 76 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 350 350 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 521 521 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 574 574 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 609 623 FVAIYGQIHTLNNDY -> SSPESCKNYQGGYY (in isoform 2).  
 FT /FTId=VSP\_001419.  
 FT CONFLICT 149 149 MISSING (IN REF. 1).  
 FT CONFLICT 220 220 S -> T (IN REF. 1).  
 FT CONFLICT 494 494 E -> D (IN REF. 1).  
 SQ SEQUENCE 623 AA; 71697 MW; AFD2724D5C1F56C8 CRC64;

Query Match 12.4%; Score 216.5; DB 1; Length 623;  
 Best Local Similarity 26.1%; Pred. No. 8.2e-10;  
 Matches 81; Conservative 41; Mismatches 107; Indels 81; Gaps 15;

Qy 32 AYKAPVVVEG-----KVQGLVPAGGSSSNSTREPPA 62  
 |: || |:| :: || : ||| :  
 Db 329 AFAAPTQVFGVFKSMSADRRVNFSA TMKVEKVYKQ QHDLQLPTLVRLQFALSNSSGECD- 387  
 Qy 63 SGRVALVKVLDKWPLRSGG-LQREQVISVGSCVPLERNQRYIFFLEPTE-----QP- 112  
 : :: : ||| ||: || |: |:| ||  
 Db 388 ---IYRERLMFRGMLRSGNDLQQASDIS-----YMMFVQQTNPNGNFTILGQPM 432  
 Qy 113 ----LVFKTAFAPLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAA 168  
 || : : | ||| ||| ::| :| |:| : ||  
 Db 433 RVTHLVVEAVETAVSEN-YTQNAEVTKIF-----SKPSKAIKH-----GKKLRIVCE-V 480  
 Qy 169 AGNPQPSYRWFKDGKELNRSRDIRIKYGNRKNRSL---QFNKVKVEDAGEYVCEAENIL 225  
 :| | | ||| | :|| |:| :: : : | | || ||| | | :|  
 Db 481 SGQPPPKVTWFKDEKSINRKRNI-YQFKHHKRRSELIVRSFN--SSSDAGRYECRAKNKA 537  
 Qy 226 GKDTVGRRLYVNSVSTTL-SSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGF 284  
 | : |: : : : | || || : | | | ||| : |:| :  
 Db 538 SKAIAKRIRIMIKASPVHFPTDRSASGIPCN---FDYCFHNGTCRMIPDINEVYCRCPTEY 594

Qy 285 FGQRCLEKLP 294  
|| || | |  
Db 595 FGNRCEKWP 604

RESULT 9

JAM2\_HUMAN

ID JAM2\_HUMAN STANDARD; PRT; 298 AA.  
AC P57087;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Junctional adhesion molecule 2 precursor (Vascular endothelial  
DE junction-associated molecule) (VE-JAM).  
GN JAM2 OR VEJAM OR C21ORF43.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Vascular endothelial cells;  
RX MEDLINE=20317114; PubMed=10779521;  
RA Palmeri D., van Zante A., Huang C.C., Hemmerich S., Rosen S.D.;  
RT "Vascular endothelial junction-associated molecule, a novel member of  
RT the immunoglobulin superfamily, is localized to intercellular  
RT boundaries of endothelial cells.";  
RL J. Biol. Chem. 275:19139-19145(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=20507930; PubMed=10945976;  
RA Cunningham S.A., Arrate M.P., Rodriguez J.M., Bjercke R.J.,  
RA Vanderslice P., Morris A.P., Brock T.A.;  
RT "A novel protein with homology to the junctional adhesion molecule:  
RT Characterization of leukocyte interactions.";  
RL J. Biol. Chem. 275:34750-34756(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,



RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- FUNCTION: MAY PLAY A ROLE IN THE PROCESSES OF LYMPHOCYTE HOMING TO  
 CC SECONDARY LYMPHOID ORGANS.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).  
 CC -!- TISSUE SPECIFICITY: PROMINENTLY EXPRESSED ON HIGH ENDOTHELIAL  
 CC VENULES BUT IS ALSO PRESENT ON THE ENDOTHELIA OF OTHER VESSELS.  
 CC LOCALIZED TO THE INTERCELLULAR BOUNDARIES OF HIGH ENDOTHELIAL  
 CC CELLS.  
 CC -!- SIMILARITY: Belongs to the immunoglobulin superfamily.  
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.  
 CC -!- DATABASE: NAME=PROW; NOTE=PROW 2:1-3(2001);  
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/guide/1652492186\_g.htm".  
 CC

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AF255910; AAF81223.1; -.  
 DR EMBL; AY016009; AAG49022.1; -.  
 DR EMBL; BC017779; AAH17779.1; -.  
 DR Genew; HGNC:14686; JAM2.  
 DR MIM; 606870; -.  
 DR GO; GO:0005887; C:integral to plasma membrane; NAS.  
 DR GO; GO:0016337; P:cell-cell adhesion; NAS.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR Pfam; PF00047; ig; 2.  
 DR SMART; SM00408; IGc2; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 2.  
 KW Immunoglobulin domain; Glycoprotein; Transmembrane; Signal.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 298 JUNCTIONAL ADHESION MOLECULE 2.  
 FT DOMAIN 21 238 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 239 259 POTENTIAL.  
 FT DOMAIN 260 298 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 32 127 IG-LIKE V-TYPE.  
 FT DOMAIN 134 238 IG-LIKE C2-TYPE.  
 FT DISULFID 50 109 POTENTIAL.  
 FT DISULFID 155 214 POTENTIAL.  
 FT CARBOHYD 98 98 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 236 236 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 298 AA; 33207 MW; CA78E518E22DCAEE CRC64;

Query Match 8.9%; Score 155; DB 1; Length 298;  
 Best Local Similarity 27.7%; Pred. No. 2.9e-05;  
 Matches 56; Conservative 24; Mismatches 86; Indels 36; Gaps 8;

```

ID      LAMP_RAT          STANDARD;          PRT;      338 AA.
AC      Q62813;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Limbic system-associated membrane protein precursor (LSAMP).
GN      LSAMP OR LAMP.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A., AND SEQUENCE OF 29-49.
RC      TISSUE=Hippocampus;
RX      MEDLINE=95374785; PubMed=7646886;
RA      Pimenta A.F., Zhukareva V., Barbe M.F., Reinoso B.S., Grimley C.,
RA      Henzel W., Fischer I., Levitt P.;
RT      "The limbic system-associated membrane protein is an Ig superfamily
RT      member that mediates selective neuronal growth and axon targeting.";
RL      Neuron 15:287-297(1995).
CC      -!- FUNCTION: MEDIATES SELECTIVE NEURONAL GROWTH AND AXON TARGETING.
CC      CONTRIBUTES TO THE GUIDANCE OF DEVELOPING AXONS AND REMODELING OF
CC      MATURE CIRCUITS IN THE LIMBIC SYSTEM. ESSENTIAL FOR NORMAL GROWTH
CC      OF THE HYPPOCAMPAL MOSSY FIBER PROJECTION.
CC      -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC      -!- TISSUE SPECIFICITY: Expressed mostly by neurons comprising limbic-
CC      associated cortical and subcortical regions that function in
CC      cognition, emotion, memory, and learning.
CC      -!- DEVELOPMENTAL STAGE: FIRST DETECTED AT E15-16, AT STAGE E20 IT IS
CC      DETECTED IN PRESUMPTIVE CORTEX, MEDIAL LIMBIC AREAS OF THE
CC      THALAMUS AND HYPOTHALAMUS. IN THE ADULT, IT IS FOUND IN
CC      HYPOTHALAMUS, PERIRHINAL CORTEX, AMYGDALA AND MEDIAL THALAMIC
CC      REGION.
CC      -!- SIMILARITY: Belongs to the immunoglobulin superfamily. IgLON
CC      family.
CC      -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 DR EMBL; U31554; AAA86120.1; -.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR Pfam; PF00047; ig; 3.  
 DR SMART; SM00408; IGc2; 2.  
 DR PROSITE; PS50835; IG\_LIKE; 3.  
 KW Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;  
 KW Repeat; Signal; Lipoprotein.  
 FT SIGNAL 1 28  
 FT CHAIN 29 315 LIMBIC SYSTEM-ASSOCIATED MEMBRANE  
 FT PROTEIN.  
 FT PROPEP 316 338 REMOVED IN MATURE FORM (POTENTIAL).  
 FT DOMAIN 29 122 IG-LIKE C2-TYPE 1.  
 FT DOMAIN 132 214 IG-LIKE C2-TYPE 2.  
 FT DOMAIN 219 304 IG-LIKE C2-TYPE 3.  
 FT DISULFID 53 111 POTENTIAL.  
 FT DISULFID 153 197 POTENTIAL.  
 FT DISULFID 239 290 POTENTIAL.  
 FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 279 279 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 315 315 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT LIPID 315 315 GPI-anchor amidated asparagine  
 FT (Potential).  
 SQ SEQUENCE 338 AA; 37324 MW; 0B76AFDD68A39BB6 CRC64;

Query Match 8.3%; Score 145; DB 1; Length 338;  
 Best Local Similarity 23.9%; Pred. No. 0.00021;  
 Matches 55; Conservative 31; Mismatches 92; Indels 52; Gaps 9;

Qy 53 SSNSTREPPASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQP 112  
 | : || | :|| | : : : : || | : :  
 Db 112 SVQTQHEPKTSQVYLIVQV----PPKISNISSDVTVNEGSNVTL-----VCMANGRPEP 161  
 Qy 113 LVFKTAFAPL-----DTNGKNLKKEVGKILCTDCAT-----RPKL 147  
 :: || : :|| | : : | :  
 Db 162 VITWRHLTPLGREFEGEEYLEILGITREQSGKYECKAANEVSSADVKQVKVTVNYPPTI 221  
 Qy 148 KMKKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRSRDIRIKYGNRKNRSLQFN 207  
 : || | : ||||| : | | : :|| : : : || : |  
 Db 222 TESKSNEATTGRQASLKCEASA-VPAPDFEYRDDTRINSANGLEIKSTEGQ--SSLTVT 278  
 Qy 208 KVKVEDAGEYVCEAENILG-----KDTVGRGRLYVN-SVSTTLSSW 246  
 | | | | | | :|| :| :| : |  
 Db 279 NVTEEHYGNITCVAANKLGVTNASLVLFPGSVRG---INGSISLAVPLW 325

RESULT 11

LAMP\_HUMAN

ID LAMP\_HUMAN STANDARD; PRT; 338 AA.  
 AC Q13449;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Limbic system-associated membrane protein precursor (LSAMP).  
 GN LSAMP OR LAMP.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96235133; PubMed=8666243;  
 RA Pimenta A.F., Fischer I., Levitt P.;  
 RT "cDNA cloning and structural analysis of the human limbic-system-  
 RT associated membrane protein (LAMP).";  
 RL Gene 170:189-195(1996).

CC -!- FUNCTION: MEDIATES SELECTIVE NEURONAL GROWTH AND AXON TARGETING.  
 CC CONTRIBUTES TO THE GUIDANCE OF DEVELOPING AXONS AND REMODELING OF  
 CC MATURE CIRCUITS IN THE LIMBIC SYSTEM. ESSENTIAL FOR NORMAL GROWTH  
 CC OF THE HYPOCAMPAL MOSSY FIBER PROJECTION (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -!- TISSUE SPECIFICITY: Expressed on limbic neurons and fiber tracts  
 CC as well as in single layers of the superior colliculus, spinal  
 CC chord and cerebellum.  
 CC -!- SIMILARITY: Belongs to the immunoglobulin superfamily. IgLON  
 CC family.  
 CC -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; U41901; AAC50569.1; -.  
 DR PIR; JC4776; JC4776.  
 DR Genew; HGNC:6705; LSAMP.  
 DR MIM; 603241; -.  
 DR GO; GO:0007399; P:neurogenesis; TAS.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR Pfam; PF00047; ig; 3.  
 DR SMART; SM00408; IGc2; 2.  
 DR PROSITE; PS50835; IG\_LIKE; 3.  
 KW Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;  
 KW Repeat; Signal; Lipoprotein.  
 FT SIGNAL 1 28 POTENTIAL.  
 FT CHAIN 29 315 LIMBIC SYSTEM-ASSOCIATED MEMBRANE  
 FT PROTEIN.  
 FT PROPEP 316 338 REMOVED IN MATURE FORM (POTENTIAL).

FT	DOMAIN	29	122	IG-LIKE C2-TYPE 1.
FT	DOMAIN	132	214	IG-LIKE C2-TYPE 2.
FT	DOMAIN	219	304	IG-LIKE C2-TYPE 3.
FT	DISULFID	53	111	POTENTIAL.
FT	DISULFID	153	197	POTENTIAL.
FT	DISULFID	239	290	POTENTIAL.
FT	CARBOHYD	40	40	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	66	66	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	136	136	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	148	148	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	279	279	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	287	287	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	300	300	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	315	315	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	LIPID	315	315	GPI-anchor amidated asparagine
FT				(Potential).
SQ	SEQUENCE	338 AA;	37308 MW;	03455F286DF5D92F CRC64;

Query Match 8.2%; Score 144; DB 1; Length 338;  
 Best Local Similarity 25.0%; Pred. No. 0.00026;  
 Matches 56; Conservative 34; Mismatches 94; Indels 40; Gaps 10;

Qy	53	SSNSTREPPASGRVALVKVL	DKWPLRSGGLQREQVISVGS-----CVPLERNQRYIFF--	105
		:      :     : : :     :   :   :		
Db	112	SVQTQHEPKTSQVYLIVQV----	PPKISNISSDVTVNEGSNVTLCMANGRPEPVITWRH	167
Qy	106	LEPTEQPLVFKTAFAPL-----	DTNGKNLKEVGKILCTDCAT-----RPKLKKMKSQ	153
		: : : : :     : :     : :		
Db	168	LTPTGREFEGEEYLEILGITRE	QSGKYECKAANEVSSADVQVKVTVNYPPTITESKSN	227
Qy	154	TGQVGEKQSLKCEAAAGNPQ	PSYRWFKDGGKELNRSRDIRIKYGNRKN	SRLQFNKVKVED 213
		:       :    :    : :   :   :		
Db	228	EATTGRQASLKCEASA-VPAP	DFEYRDDTRINSANGLEIKSTEGQ--SSLTVTNVTEEH	284
Qy	214	AGEYVCEAENILG-----	KDTRVGRRLYN-SVSTTLSSW	246
		:   :    :		
Db	285	YGNITCVAANKLGVTNASLV	LFPGSVRG---INGSISLAVPLW	325

# RESULT 12

## CEPU\_CHICK

ID CEPU\_CHICK STANDARD; PRT; 353 AA.  
 AC Q90773;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE CEPU-1 protein precursor.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC TISSUE=Brain;  
 RX MEDLINE=96370549; PubMed=8774445;  
 RA Spaltmann F., Bruemmendorf T.;

```

RT  "CEPU-1, a novel immunoglobulin superfamily molecule, is expressed by
RT  developing cerebellar Purkinje cells.";
RL  J. Neurosci. 16:1770-1779(1996).
CC  -!- FUNCTION: It may be a cellular address molecule specific to
CC      Purkinje cells. It may represent a receptor or a subunit of a
CC      receptor complex.
CC  -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC  -!- ALTERNATIVE PRODUCTS:
CC      Event=Alternative splicing; Named isoforms=2;
CC      Name=1; Synonyms=Minor;
CC      IsoId=Q90773-1; Sequence=Displayed;
CC      Name=2; Synonyms=Major;
CC      IsoId=Q90773-2; Sequence=VSP_002607;
CC  -!- TISSUE SPECIFICITY: Found on the dendrites, somata and axons of
CC      developing Purkinje cells. Undetectable on other neurons like
CC      Golgi or granule cells.
CC  -!- DEVELOPMENTAL STAGE: Expressed by developing cerebellar Purkinje
CC      cells. Expression coincides with the growth of the dendritic tree,
CC      after Purkinje cells have finished their migration from the
CC      ventricular zone (from E15 until E21). Expressed in the adult.
CC  -!- SIMILARITY: Belongs to the immunoglobulin superfamily. IgLON
CC      family.
CC  -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; Z72497; CAA96578.1; -.
DR  InterPro; IPR007110; Ig-like.
DR  InterPro; IPR003598; Ig_c2.
DR  Pfam; PF00047; ig; 3.
DR  SMART; SM00408; IGc2; 2.
DR  PROSITE; PS50835; IG_LIKE; 3.
KW  Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW  Repeat; Signal; Alternative splicing; Lipoprotein.
FT  SIGNAL      1      28      POTENTIAL.
FT  CHAIN       29     330      CEPU-1 PROTEIN.
FT  PROPEP     331     353      REMOVED IN MATURE FORM (POTENTIAL).
FT  DOMAIN      37     124      IG-LIKE C2-TYPE 1.
FT  DOMAIN     134     216      IG-LIKE C2-TYPE 2.
FT  DOMAIN     220     314      IG-LIKE C2-TYPE 3.
FT  DISULFID    55     113      POTENTIAL.
FT  DISULFID   155     199      POTENTIAL.
FT  DISULFID   241     293      POTENTIAL.
FT  CARBOHYD    42      42      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD    68      68      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD   150     150      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD   282     282      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD   290     290      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD   303     303      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  LIPID       330     330      GPI-anchor amidated serine (Potential).
FT  VARSPLIC   310     320      Missing (in isoform 2).

```

FT /FTId=VSP\_002607.  
SQ SEQUENCE 353 AA; 38736 MW; 2550C48591EBBBA6 CRC64;

Query Match 8.1%; Score 141.5; DB 1; Length 353;  
Best Local Similarity 36.5%; Pred. No. 0.00043;  
Matches 38; Conservative 11; Mismatches 52; Indels 3; Gaps 3;

Qy 145 PKLKMKMQTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRSRDIRIKYGNRKN SRL 204  
| : || ||:| | ||:| | :|:| | | : :| | ||  
Db 221 PYISDAKSTGVPVGQKGILMCEASA-VPSADFQWKDDKRLAEGQK-GLKVENKAFFSRL 278  
  
Qy 205 QFNKVKVEDAGEYVCEAENILGKDTVGRLYVNSVSTTLSSWSG 248  
| | :| | | | | | | | | :| | :| | |  
Db 279 TFFNVSEQDYGNITCVASNQLGNTNASMILY-EETTTALTTPWKG 321

### RESULT 13

#### NCA2\_HUMAN

ID NCA2\_HUMAN STANDARD; PRT; 761 AA.  
AC P13592; P13593;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Neural cell adhesion molecule 1, 120 kDa isoform precursor (N-CAM 120)  
DE (NCAM-120) (CD56 antigen).  
GN NCAM1 OR NCAM.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM N-CAM 120).  
RC TISSUE=Skeletal muscle;  
RX MEDLINE=89305258; PubMed=3253057;  
RA Barton C.H., Dickson G., Gower H.J., Rowett L.H., Putt W.,  
RA Elsom V., Moore S.E., Goridis C., Walsh F.S.;  
RT "Complete sequence and in vitro expression of a tissue-specific  
RT phosphatidylinositol-linked N-CAM isoform from skeletal muscle."  
RL Development 104:165-173(1988).  
RN [2]  
RP SEQUENCE OF 491-761 FROM N.A. (ISOFORM N-CAM 120).  
RC TISSUE=Skeletal muscle;  
RX MEDLINE=87301755; PubMed=2887295;  
RA Dickson G., Gower H.J., Barton C.H., Prentice H.M., Elsom V.L.,  
RA Moore S.E., Cox R.D., Quinn C., Putt W., Walsh F.S.;  
RT "Human muscle neural cell adhesion molecule (N-CAM): identification  
RT of a muscle-specific sequence in the extracellular domain."  
RL Cell 50:1119-1130(1987).  
RN [3]  
RP SEQUENCE OF 491-655 FROM N.A. (ISOFORM C).  
RX MEDLINE=89077552; PubMed=3203385;  
RA Gower H.J., Barton C.H., Elsom V.L., Thompson J., Moore S.E.,  
RA Dickson G., Walsh F.S.;  
RT "Alternative splicing generates a secreted form of N-CAM in muscle  
RT and brain."  
RL Cell 55:955-964(1988).  
CC -!- FUNCTION: This protein is a cell adhesion molecule involved in

```

CC      neuron-neuron adhesion, neurite fasciculation, outgrowth of
CC      neurites, etc.
CC      -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC      -!- ALTERNATIVE PRODUCTS:
CC          Event=Alternative splicing; Named isoforms=3;
CC          Name=N-CAM 120;
CC              IsoId=P13592-2; Sequence=Displayed;
CC          Name=N-CAM 140;
CC              IsoId=P13591-1; Sequence=External;
CC          Name=C; Synonyms=Secreted;
CC              IsoId=P13592-1; Sequence=VSP_002587;
CC      -!- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.
CC      -!- SIMILARITY: Contains 2 fibronectin type III domains.
CC      -!- DATABASE: NAME=PROW; NOTE=CD guide CD56 entry;
CC          WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd56.htm".
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X16841; CAA34739.1; -.
DR      EMBL; M17409; AAA59912.1; -.
DR      EMBL; M22094; AAA59910.1; -.
DR      EMBL; M22092; AAA59911.1; -.
DR      EMBL; M22091; AAA59911.1; JOINED.
DR      PIR; A31635; A31635.
DR      PIR; S07784; IJHUNG.
DR      Genew; HGNC:7656; NCAM1.
DR      MIM; 116930; -.
DR      GO; GO:0016021; C:integral to membrane; TAS.
DR      GO; GO:0005886; C:plasma membrane; TAS.
DR      InterPro; IPR008957; FN_III-like.
DR      InterPro; IPR003961; FN_III.
DR      InterPro; IPR007110; Ig-like.
DR      InterPro; IPR003598; Ig_c2.
DR      Pfam; PF00041; fn3; 2.
DR      Pfam; PF00047; ig; 5.
DR      SMART; SM00060; FN3; 2.
DR      SMART; SM00408; IGc2; 5.
DR      PROSITE; PS50835; IG_LIKE; 5.
KW      Immunoglobulin domain; Cell adhesion; Glycoprotein; Repeat; Signal;
KW      GPI-anchor; Alternative splicing.
FT      SIGNAL      1      19
FT      CHAIN      20      761      NEURAL CELL ADHESION MOLECULE 1, 120 kDa
FT                                     ISOFORM.
FT      DOMAIN      20      111      IG-LIKE C2-TYPE 1.
FT      DOMAIN      116     205      IG-LIKE C2-TYPE 2.
FT      DOMAIN      212     301      IG-LIKE C2-TYPE 3.
FT      DOMAIN      308     403      IG-LIKE C2-TYPE 4.
FT      DOMAIN      406     491      IG-LIKE C2-TYPE 5.
FT      DOMAIN      518     595      FIBRONECTIN TYPE-III 1.
FT      DOMAIN      660     727      FIBRONECTIN TYPE-III 2.
FT      DISULFID     41      96      PROBABLE.

```



FT	DISULFID	139	189	PROBABLE.
FT	DISULFID	235	287	PROBABLE.
FT	DISULFID	329	385	PROBABLE.
FT	DISULFID	426	479	PROBABLE.
FT	CARBOHYD	222	222	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	315	315	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	347	347	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	423	423	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	449	449	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	478	478	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	635	655	GEPSAPKLEGQMGEDGNSIKV -> NIAQNHCCNMFFQAGLH
FT				NALMK (in isoform C).
FT				/FTid=VSP_002587.
SQ	SEQUENCE	761 AA;	83770 MW;	F0CAD3292D7AB67E CRC64:

Qy	154	TGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRSRDIRIKYGNGRKNSRLQFNKVKVED	213
		:  :   :      :  :       :  : :       :	
Db	224	TANLGQSVTLVCD-AEGFPEPTMSWTKDGEQIEQEEDDE-KYIFSDDSSQLTIKKVDKND	281
Qy	214	AGEYVCEAENILGKD--TVRGRLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYYIE	271
		:        :   : :  :   :       :    :	
Db	282	EAEYICIAENKAGEQDATIHLKVFAKPKITYVE-----NQTAME-----LE	322
Qy	272	GINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQSVLWDTPTGTGVSSSQWSTSPSTLD	328
		:  :               :    :	
Db	323	EQVTLTCEASG-----DPIPSITWRTSTRNISSEE-----KTLD	356

## RESULT 14

```

ID      NCAL_HUMAN          STANDARD;          PRT;      848 AA.
AC      P13591; Q15829; Q16180;
DT      01-JAN-1990 (Rel. 13, Created)
DT      15-JUL-1998 (Rel. 36, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Neural cell adhesion molecule 1, 140 kDa isoform precursor (N-CAM 140)
DE      (NCAM-140) (CD56 antigen).
GN      NCAM1 OR NCAM.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=94356433; PubMed=8075973;
RA      Saito S., Tanio Y., Tachibana I., Hayashi S., Kishimoto T., Kawase I.;
RT      "Complementary DNA sequence encoding the major neural cell adhesion
RT      molecule isoform in a human small cell lung cancer cell line.";
RL      Lung Cancer 10:307-318(1994).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=91250739; PubMed=1710251;
RA      Lanier L.L., Chang C., Azuma M., Ruitenberg J.J., Hemperly J.J.,

```

RA Phillips J.H.;  
 RT "Molecular and functional analysis of human natural killer cell-  
 RT associated neural cell adhesion molecule (N-CAM/CD56).";  
 RL J. Immunol. 146:4421-4426(1991).  
 RN [3]  
 RP SEQUENCE OF 491-848 FROM N.A.  
 RX MEDLINE=87301755; PubMed=2887295;  
 RA Dickson G., Gower H.J., Barton C.H., Prentice H.M., Elsom V.L.,  
 RA Moore S.E., Cox R.D., Quinn C., Putt W., Walsh F.S.;  
 RT "Human muscle neural cell adhesion molecule (N-CAM): identification  
 RT of a muscle-specific sequence in the extracellular domain.";  
 RL Cell 50:1119-1130(1987).  
 CC -!- FUNCTION: This protein is a cell adhesion molecule involved in  
 CC neuron-neuron adhesion, neurite fasciculation, outgrowth of  
 CC neurites, etc.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Name=N-CAM 140;  
 CC IsoId=P13591-1; Sequence=Displayed;  
 CC Name=N-CAM 120;  
 CC IsoId=P13592-2; Sequence=External;  
 CC Name=C; Synonyms=Secreted;  
 CC IsoId=P13592-1; Sequence=External;  
 CC -!- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.  
 CC -!- SIMILARITY: Contains 2 fibronectin type III domains.  
 CC -!- DATABASE: NAME=PROW; NOTE=CD guide CD56 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd56.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; S71824; AAB31836.1; -.  
 DR EMBL; U63041; AAB04558.1; -.  
 DR EMBL; M17410; AAA59913.1; -.  
 DR HSSP; P40189; 1BQU.  
 DR Genew; HGNC:7656; NCAM1.  
 DR MIM; 116930; -.  
 DR GO; GO:0016021; C:integral to membrane; TAS.  
 DR GO; GO:0005886; C:plasma membrane; TAS.  
 DR InterPro; IPR008957; FN\_III-like.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR Pfam; PF00041; fn3; 2.  
 DR Pfam; PF00047; ig; 5.  
 DR SMART; SM00060; FN3; 2.  
 DR SMART; SM00408; IGc2; 5.  
 DR PROSITE; PS50835; IG\_LIKE; 5.  
 KW Immunoglobulin domain; Cell adhesion; Glycoprotein; Repeat; Signal;  
 KW Transmembrane; Alternative splicing.  
 FT SIGNAL 1 19

FT	CHAIN	20	848	NEURAL CELL ADHESION MOLECULE 1, 140 kDa
FT				ISOFORM.
FT	DOMAIN	20	708	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	709	729	POTENTIAL.
FT	DOMAIN	730	848	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	20	111	IG-LIKE C2-TYPE 1.
FT	DOMAIN	116	205	IG-LIKE C2-TYPE 2.
FT	DOMAIN	212	301	IG-LIKE C2-TYPE 3.
FT	DOMAIN	308	403	IG-LIKE C2-TYPE 4.
FT	DOMAIN	406	491	IG-LIKE C2-TYPE 5.
FT	DOMAIN	518	595	FIBRONECTIN TYPE-III 1.
FT	DOMAIN	660	727	FIBRONECTIN TYPE-III 2.
FT	DISULFID	41	96	PROBABLE.
FT	DISULFID	139	189	PROBABLE.
FT	DISULFID	235	287	PROBABLE.
FT	DISULFID	329	385	PROBABLE.
FT	DISULFID	426	479	PROBABLE.
FT	CARBOHYD	222	222	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	315	315	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	347	347	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	423	423	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	449	449	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	478	478	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	215	215	Q -> R (IN REF. 2).
FT	CONFLICT	239	239	G -> R (IN REF. 2).
FT	CONFLICT	490	490	L -> F (IN REF. 2).
FT	CONFLICT	599	600	QG -> R (IN REF. 3).
FT	CONFLICT	720	721	MISSING (IN REF. 3).
FT	CONFLICT	811	811	G -> A (IN REF. 3).
SQ	SEQUENCE	848 AA;	93360 MW;	68D2F0C0E6C1C2AD CRC64;

Query Match 8.0%; Score 140; DB 1; Length 848;

Best Local Similarity 25.4%; Pred. No. 0.0016;

Matches 45; Conservative 26; Mismatches 60; Indels 46; Gaps 7;

Qy	154	TGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNGRKNSRLQFNKVKVED	213
		: : :   :      : :      :: :      : :	
Db	224	TANLGQSVTLVCD-AEGFPEPTMSWTKDGEQIEQEEDDE-KYIFSDSSQLTIKKVDKND	281
Qy	214	AGEYVCEAENILGKD--TVRGRLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCIYIE	271
		:       :  : ::   :  :  :	
Db	282	EAEYICIAENKAGEQDATIHLKVFAKPKITYVE-----NQATME-----LE	322
Qy	272	GINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQSVLWDTPTGTGVSSSQWSTSPSTLD	328
		: :     :     :  :	
Db	323	EQVTLTCEASG-----DPIPSITWRTSTRNISSEE-----KTLD	356

RESULT 15

UN89 CAEEL

ID UN89 CAEEL STANDARD; PRT; 6632 AA.

AC 001761; Q17362;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).

GN    UNC-89   OR   C09D1.1.

OS *Caenorhabditis elegans*.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; *Caenorhabditis*.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=96180278; PubMed=8603916;  
 RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;  
 RT "The *Caenorhabditis elegans* gene *unc-89*, required for muscle M-line  
 RT assembly, encodes a giant modular protein composed of Ig and signal  
 RT transduction domains.";  
 RL J. Cell Biol. 132:835-848(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Du Z., Le T.T., Wilson R.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP REVISIONS.  
 RA Waterston R.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: Structural component of the muscle M-line. Myofilament  
 CC lattice assembly begins with positional cues laid down in the  
 CC basement membrane and muscle cell membrane. UNC-89 responds to  
 CC these signals, localizes, and then participates in assembling an  
 CC M-line.  
 CC -!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.  
 CC -!- SIMILARITY: Contains 1 DBL-homology (DH) domain.  
 CC -!- SIMILARITY: Contains 1 fibronectin type III domain.  
 CC -!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.  
 CC -!- SIMILARITY: Contains 1 PH domain.  
 CC -!- SIMILARITY: Contains 5 RCSD domains.  
 CC -!- SIMILARITY: Contains 1 SH3 domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U33058; AAB00542.1; -.  
 DR EMBL; AF003131; AAB54132.2; -.  
 DR PDB; 1FHO; 20-DEC-00.  
 DR WormPep; C09D1.1; CE30426.  
 DR InterPro; IPR008957; FN\_III-like.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR InterPro; IPR001849; PH.  
 DR InterPro; IPR007850; RCSD.  
 DR InterPro; IPR000219; RhoGEF.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00041; fn3; 1.

DR Pfam; PF00047; ig; 47.  
 DR Pfam; PF00169; PH; 1.  
 DR Pfam; PF05177; RCSL; 5.  
 DR Pfam; PF00621; RhoGEF; 1.  
 DR Pfam; PF00018; SH3; 1.  
 DR SMART; SM00408; IGc2; 23.  
 DR SMART; SM00325; RhoGEF; 1.  
 DR SMART; SM00326; SH3; 1.  
 DR PROSITE; PS50010; DH\_2; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 49.  
 DR PROSITE; PS50003; PH\_DOMAIN; 1.  
 DR PROSITE; PS50002; SH3; 1.  
 KW Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;  
 KW 3D-structure.

FT	DOMAIN	63	127	SH3.
FT	DOMAIN	152	330	DH.
FT	DOMAIN	342	498	PH.
FT	DOMAIN	547	633	IG-LIKE C2-TYPE 1.
FT	DOMAIN	648	736	IG-LIKE C2-TYPE 2.
FT	DOMAIN	748	838	IG-LIKE C2-TYPE 3.
FT	DOMAIN	946	1033	IG-LIKE C2-TYPE 4.
FT	DOMAIN	1044	1132	IG-LIKE C2-TYPE 5.
FT	DOMAIN	1140	1227	IG-LIKE C2-TYPE 6.
FT	DOMAIN	1272	1315	THR-RICH.
FT	DOMAIN	1375	1475	RCSL 1.
FT	DOMAIN	1479	1585	RCSL 2.
FT	DOMAIN	1597	1695	RCSL 3.
FT	DOMAIN	1700	1799	RCSL 4.
FT	DOMAIN	1800	1860	RCSL 5.
FT	DOMAIN	1982	2067	IG-LIKE C2-TYPE 7.
FT	DOMAIN	2071	2163	IG-LIKE C2-TYPE 8.
FT	DOMAIN	2171	2261	IG-LIKE C2-TYPE 9.
FT	DOMAIN	2269	2359	IG-LIKE C2-TYPE 10.
FT	DOMAIN	2367	2455	IG-LIKE C2-TYPE 11.
FT	DOMAIN	2463	2564	IG-LIKE C2-TYPE 12.
FT	DOMAIN	2563	2651	IG-LIKE C2-TYPE 13.
FT	DOMAIN	2657	2746	IG-LIKE C2-TYPE 14.
FT	DOMAIN	2754	2858	IG-LIKE C2-TYPE 15.
FT	DOMAIN	2887	2980	IG-LIKE C2-TYPE 16.
FT	DOMAIN	2994	3081	IG-LIKE C2-TYPE 17.
FT	DOMAIN	3087	3183	IG-LIKE C2-TYPE 18.
FT	DOMAIN	3189	3280	IG-LIKE C2-TYPE 19.
FT	DOMAIN	3286	3376	IG-LIKE C2-TYPE 20.
FT	DOMAIN	3384	3469	IG-LIKE C2-TYPE 21.
FT	DOMAIN	3482	3572	IG-LIKE C2-TYPE 22.
FT	DOMAIN	3580	3667	IG-LIKE C2-TYPE 23.
FT	DOMAIN	3686	3777	IG-LIKE C2-TYPE 24.
FT	DOMAIN	3817	3908	IG-LIKE C2-TYPE 25.
FT	DOMAIN	3920	4009	IG-LIKE C2-TYPE 26.
FT	DOMAIN	4018	4106	IG-LIKE C2-TYPE 27.
FT	DOMAIN	4109	4201	IG-LIKE C2-TYPE 28.
FT	DOMAIN	4212	4297	IG-LIKE C2-TYPE 29.
FT	DOMAIN	4302	4387	IG-LIKE C2-TYPE 30.
FT	DOMAIN	4400	4485	IG-LIKE C2-TYPE 31.
FT	DOMAIN	4489	4580	IG-LIKE C2-TYPE 32.
FT	DOMAIN	4588	4678	IG-LIKE C2-TYPE 33.
FT	DOMAIN	4681	4771	IG-LIKE C2-TYPE 34.

FT	DOMAIN	4873	4961	IG-LIKE C2-TYPE 35.
FT	DOMAIN	4965	5057	IG-LIKE C2-TYPE 36.
FT	DOMAIN	5067	5160	IG-LIKE C2-TYPE 37.
FT	DOMAIN	5171	5260	IG-LIKE C2-TYPE 38.
FT	DOMAIN	5277	5366	IG-LIKE C2-TYPE 39.
FT	DOMAIN	5383	5472	IG-LIKE C2-TYPE 40.
FT	DOMAIN	5487	5578	IG-LIKE C2-TYPE 41.
FT	DOMAIN	5595	5685	IG-LIKE C2-TYPE 42.
FT	DOMAIN	5701	5790	IG-LIKE C2-TYPE 43.
FT	DOMAIN	5815	5904	IG-LIKE C2-TYPE 44.
FT	DOMAIN	5925	6014	IG-LIKE C2-TYPE 45.
FT	DOMAIN	6038	6130	IG-LIKE C2-TYPE 46.
FT	DOMAIN	6150	6239	IG-LIKE C2-TYPE 47.
FT	DOMAIN	6275	6368	FIBRONECTIN TYPE-III.
FT	DOMAIN	6413	6502	IG-LIKE C2-TYPE 48.
FT	DOMAIN	6507	6596	IG-LIKE C2-TYPE 49.
FT	DISULFID	568	621	POTENTIAL.
FT	DISULFID	2908	2975	POTENTIAL.
FT	DISULFID	3015	3065	POTENTIAL.
FT	DISULFID	3707	3759	POTENTIAL.
FT	DISULFID	3826	3890	POTENTIAL.
FT	DISULFID	5092	5157	POTENTIAL.
FT	DISULFID	5298	5350	POTENTIAL.
FT	DISULFID	5508	5560	POTENTIAL.
FT	DISULFID	5616	5669	POTENTIAL.
FT	DISULFID	5722	5764	POTENTIAL.
FT	DISULFID	5836	5901	POTENTIAL.
FT	DISULFID	5946	5998	POTENTIAL.
FT	DISULFID	6036	6171	POTENTIAL.
FT	DISULFID	6421	6486	POTENTIAL.
FT	CONFLICT	2137	2137	A -> P (IN REF. 1).
FT	CONFLICT	2245	2247	AKA -> PKP (IN REF. 1).
FT	CONFLICT	2258	2258	A -> P (IN REF. 1).
FT	CONFLICT	2284	2284	E -> G (IN REF. 1).
FT	CONFLICT	2297	2297	M -> I (IN REF. 1).
FT	CONFLICT	3531	3531	A -> G (IN REF. 1).
FT	CONFLICT	3884	3888	DAGEY -> RRRRI (IN REF. 1).
FT	CONFLICT	3929	3929	A -> V (IN REF. 1).
FT	CONFLICT	5134	5134	A -> P (IN REF. 1).
FT	CONFLICT	5145	5145	T -> S (IN REF. 1).
FT	CONFLICT	5185	5185	G -> A (IN REF. 1).
FT	CONFLICT	5199	5199	K -> N (IN REF. 1).
FT	CONFLICT	5202	5202	L -> F (IN REF. 1).
FT	CONFLICT	5213	5213	F -> L (IN REF. 1).
FT	CONFLICT	6178	6178	A -> G (IN REF. 1).
FT	CONFLICT	6268	6268	K -> E (IN REF. 1).
SQ	SEQUENCE	6632	AA; 731665 MW; 262D3EDD62960E89 CRC64;	

Query Match 8.0%; Score 139.5; DB 1; Length 6632;  
 Best Local Similarity 29.0%; Pred. No. 0.021;  
 Matches 54; Conservative 18; Mismatches 65; Indels 49; Gaps 8;

Qy	151	KSQTGQ-----VGEKQS--LKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKY	195
		:   :          :    :  :   :   :	
Db	3809	KTDTGGRGAPEFVELLRSCVTQKQAILKCK-VKGEPRPKIKWTKEGKEVEMSARVRAEH	3867
Qy	196	GNGRKNSRLQFNKVKVEDAGEYVCEAENILGKDTVRGRLYVNSVSTTLS---SWSGHARK	252

```

      : | : | | | | | | | | | : | | |
Db    3868 KDD-GTLTLTFDNVTQADAGEYRCEAENEYGSATTEGPIIV-----TLEGAPKIDGEAPD 3921
QY    253 CNETAKSYCVNGGVVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQSVLWDTPG 312
      : | | | : | | : | | | | | | |
Db    3922 FLQPVKPAVTVGETAVLEG--KISGK-----PKPSVKWYKNG 3957
QY    313 TGVSSS 318
      : |
Db    3958 EELKPS 3963

```

Search completed: August 17, 2004, 14:11:17  
Job time : 10.4586 secs